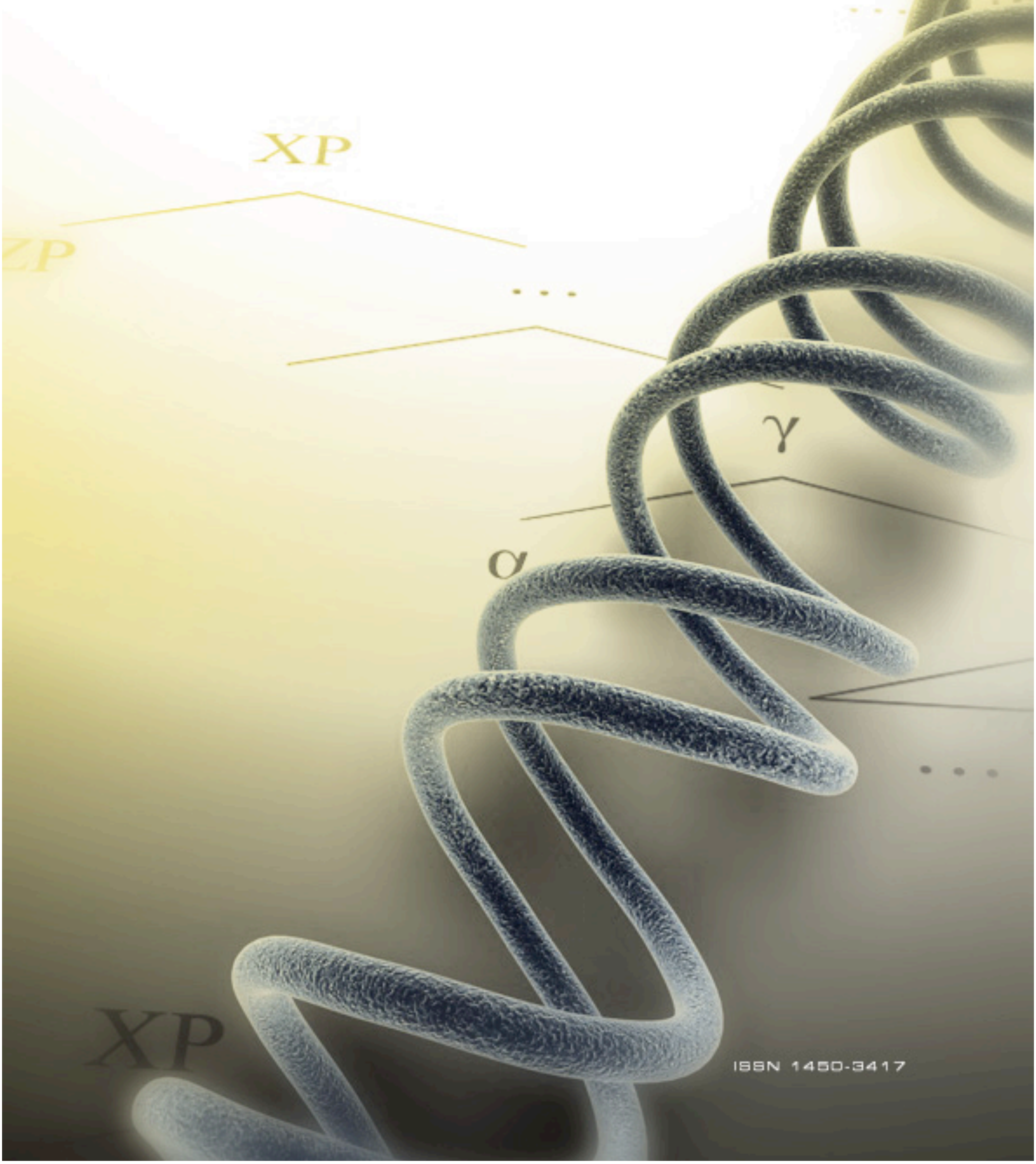
biolinguistics

volume 5
issue 3

summer
2011



ISSN 1450-3417

Space and the Vision–Language Interface: A Model-Theoretic Approach

Francesco-Alessio Ursini

The relation between spatial vision and spatial language has always been a source of controversy. Three problems can be identified as in need of a solution. A first problem pertains to the nature of the minimal information units that make up spatial vision and language. A second problem pertains to the 'dynamic' aspects of vision and language, or what visual information *to* and similar adpositions correspond to. A third problem pertains to how these different types of information are related one another, and what is the status of this 'interface', especially within a broader theory of language and cognition. The solution proposed here consists in a formal (model-theoretic) treatment of visual and linguistic information, both static and dynamic, that is couched within (a simplified form of) *Discourse Representation Theory*. It is shown that this solution is consistent with general theories of cognition and may shed some (novel) light on the nature of the FLN/FLB distinction.

Keywords: Discourse Representation Theory; faculty of language in the narrow/broad sense (FLN/FLB); interfaces; space; vision

1. Introduction: What We Talk about, When We talk about Space

In this paper, I shall address the problem of the vision–language interface: Informally, what is the exact relation between 'what we see' and 'what we say', (or: "How much space gets into language?"; Bierwisch 1996: 7). This problem can be formulated via the following (and slightly different) global research question:

Q–A: *What is the relation between vision and language?*

I shall suggest that the problem of the vision–language interface and its nature is

I wish to thank Stephen Crain, Max Coltheart, Drew Khlentzos, and Rosalind Thornton for reading and making comments on an initial draft, Qiu-Peng (Autuman) Luo and Miaoling (Anne) Hsieh for thorough feedback on later drafts, the editors of *Biolinguistics* for extensive support during the editorial process, and two anonymous reviewers for their very useful and stimulating comments, which I hope to have addressed in detail. I wish to thank the captains Harlock and Kirk as long-time role models, and my Princess for all her love and support. All remaining errors are of course my own; for those, I apologize to the reader.

biolinguistics

ISSN 1450–3417

Biolinguistics 5.3: 170–225, 2011
http://www.biolinguistics.eu

not much a problem of ‘quantity’ but ‘quality’: In order to solve this problem, we need to address not ‘how much’ information belonging to spatial representations (“what we see”) finds its way in language (and *vice versa*), but ‘how’ this process comes by and how it is possible that visual information can be realized in language in a rather flexible way. I shall argue that in order to understand how sentences such as (1) and (2) can convey non-linguistic spatial information, we need to understand how the relation between “what we see” and “what we say” comes about in the first place.

(1) Mario sits in front of the chimney.

(2) Mario has gone to the rugby match.

This problem can be solved by a *divide et impera* research strategy. I shall first split the problem in three smaller problems (the *divide* part), and solve each of them, integrating these solutions in a ‘global’ solution (the *impera* part). The three problems that constitute our central problem are the following.

First, we have a *foundational* problem, since previous proposals in the literature make different assumptions on the nature of “what we see’” and “what we say”. Some assume that language expresses only shapes of objects (as nouns) and geometrical configurations (as adpositions) (e.g., Landau & Jackendoff 1993); others that we directly express perceptual information “as we see it”, without an intermediate level of processing (i.e. language, e.g., Coventry & Garrod 2004). Hence, we don’t have a clear (theoretical) picture regarding spatial vision and spatial language, and to what extent they are distinct modules of cognition, let alone a strong, clear theory of their interface.

Second, we have a *descriptive* and *logical* problem, since previous proposals only cover inherently “static” aspects of space, but not “dynamic” aspects. Informally, these theories can account where things *are*, but not where things *are going*. Hence, we do not know what visual information adpositions such as *to* and *from* stand for, nor whether this information should be considered as “spatial’” or not.

Third, we have a *theoretical* and a *philosophica*l problem, since we must define a novel theory that is built upon the solutions to the first and second problem and can explain all the data. Then we must assess the consequences of this theory with respect to a broader theory of vision and language as part of cognition, and their unique aspects — or: What information (and properties thereof) is found in vision but not in language, and *vice versa*.

These three ‘smaller’ problems can be reformulated as the following research questions:

RQ1: *What do we know so far from the past literature, regarding spatial vision, language and their interface?*

RQ2: *What further bits of spatial knowledge must be included in our models of (spatial) vision and language, and which formal tools used to properly treat these bits?*

RQ3: *What is the nature of the vision–language interface, and which aspects are unique to language?*

Anticipating matters a bit, I shall propose the following answers. First, we know that previous literature tells us that (spatial) vision and language express internal models of objects and their possible spatial relations, and that nouns and adpositions respectively represent objects and possible relations in language. Second, we must include any type of relations in our models of vision and language, insofar as they allow establishing a relation between entities, since the emergent notion of 'space' we will obtain from our discussion is quite an abstract one. Hence, we can use a model-theoretic approach, such as *Discourse Representation Theory* (DRT; Kamp *et al*. 2005), to aptly represent these models. Third, the vision–language interface consists of the conscious processes by which we may match visual representations with linguistic ones and *vice versa*, though some linguistic representations do not represent visual objects, rather 'processes' by which we may reason about these visual objects. Consequently, vision and language can be represented as distinct models sharing the same 'logical structure', which may be connected or 'interfaced' via an opportune set of functions, representing top-down processes by which we may (consciously) evaluate whether what we see accurately describes what we say (or hear), but need not to do so.

This paper is organized as follows. In section 2, I introduce some basic notions and review previous proposals, offering an answer to the first research question. In section 3, I review theories of 'static' and 'dynamic' object recognition, and propose a model-theoretic approach to vision; I then focus on language and offer a DRT treatment of spatial language. In section 4, I integrate the two proposals in a novel theory of the vision–language interface and offer empirical evidence in support of this theory; I then focus on some of the broader consquences of the theory, by sketching an analysis of what properties emerge as unique to language from my theory, thus suggesting a somewhat novel perspective to the nature of the narrow faculty of language (FLN; Hauser *et al*. 2002, Fitch *et al*. 2005). In section 5, I finally offer my conclusions.

2. The Relation between Spatial Vision and Language

In this section I shall outline notions of spatial vision and language (section 2.1) and review previous approaches to their interface, consequently offering the first research answer (section 2.2).

2.1. Basic Notions of Space

Our daily life experiences occur in space and time,[1] as we navigate our environment by analyzing spatial relations between objects. A basic assumption, in cognitive science, is that we do so by processing (mostly) visual information about such objects and their relations as they may evolve over time, e.g., a toy which is

---

[1] Here and throughout the paper, I shall focus my attention (and use of labels) to 'space', although it would be more accurate to think of our topic as being about spatio-*temporal* vision and language, i.e. how we process location and *change* of location of objects. I hope that the lack of precision will not confuse the reader, and thank an anonymous reader for suggesting this *precís*.

on top of a table, and that we internally represent this information via a corresponding mental 'model' (e.g., Craik 1943, Johnson-Laird 1983, 1992, O'Keefe & Nadel 1978).

Another basic assumption is that, when we share this information with other fellow human beings (i.e. when we speak), we do so by defining a sub-model of space in which one object acts as the 'center' of the system, as in (3):

(3) The toy is on top of the table.

With a sentence such as (3), we convey a state of affairs in which, informally, we take the table as the origin of the reference system, take one portion of the table (its top) and assert for the toy to be more or less located in this 'area' (Talmy 1978, 2000). Our cognition of space is thus (mostly) based on the information processed and exchanged between our vision[2] module ("what we see") and our language module ("what we say"). It is also based on an emerging type of information, the structural relations that may be defined between these two modules, our ability to integrate together visual and linguistic units ("what we see and what we say") into coherent representations, over time.

The exact nature of these types of information, however, is a matter of controversy. Some say that spatial vision amounts to information about objects, their parts and shape, and the geometrical relations between these objects as when an object is on top of another (e.g., Landau & Jackendoff 1993, O'Keefe 2003). Another series of proposals offers evidence that other aspects, such as mechanical interactions (a table supporting a toy) and more abstract properties play a crucial role in how we mentally represent space (Coventry & Garrod 2004 and references therein).

We can thus observe that there is a certain tension between 'narrower', or purely geometrical, approaches and 'broader' approaches to both vision and language; as a consequence, there is also a certain tension between theories that consider spatial vision 'richer' than spatial language (e.g., Landau & Jackendoff 1993), and theories that do not assume such difference, often by simply collapsing these two modules into 'cognition' (e.g., Coventry & Garrod 2004). We thus do not have a clear picture of what information is spatial language, and what is spatial vision.

The problem of the exact type of spatial information, however, takes an even more complex nature when we look at another way in which we process spatial information, which can be loosely labeled as 'change'. Take a sentence such as (4):

(4) Mario is going to the rugby stadium.

Intuitively, this sentence describes a state of affairs in which the locatum(s) changes position over a certain amount of time of which we are aware. Mario can

---

[2] The notion of spatial vision and cognition are somewhat interchangeable for most authors. In this paper I shall use the term 'spatial vision' and 'spatial language' to avoid this confusion. Thanks to an anonymous reviewer for pointing me to this issue.

start at some unspecified starting point, move for a while, and then stop once he's at his planned destination (the rugby stadium). While there are theories of 'dynamic' vision, or how we keep track of objects changing position, as well as theories of 'dynamic' language and more specifically adpositions such as *to*, no one has attempted to integrate these theories into a broader theory of spatial vision and language, let alone in a theory of the vision–language interface.

Another challenge comes from purely linguistic facts, and what kind of information is in a sense 'unique' to a linguistic level of representation. Take a sentence such as (5):

(5) Every boy is going to a rugby field.

In this case, we can have a certain number of boys involved in the corresponding state of affairs, and each of them is described as moving in direction of a rugby field. Yet, if there are several fields at which the children can arrive (Paul goes to Manly's Oval, Joe to Randwick Field, etc.), the sentence may describe slightly different states of affairs, since they informally describe a 'collection' of more specific relations, and what they have in common. As these facts show, we need to take a broader and more flexible perspective in order to address the issue of the vision–language interface than the one usually assumed in the literature, as well as assessing in detail what elements of previous proposals we can maintain in our novel approach. Hence, I am also suggesting that the solution to this problem will offer us a quite different, but hopefully correct, answer to the 'problem of space'. Before offering this answer, however, I shall review the previous literature.

2.2. *Previous Literature*

Previous proposals on the vision–language interface can be divided into a 'narrower', geometric approach (or: "spatial language expresses geometric relations") and 'broader', 'functional' approach (or: "spatial language also expresses extra-geometrical relations"). One well-known and influential example of the geometric approach is Landau & Jackendoff (1993, henceforth L&J), while a well-known and influential functional approach is the *Functional Geometric Framework* (FGF; Coventry & Garrod 2004). I will offer a review of both, highlighting their features and shortcomings, with respect to the topic of this chapter, starting from L&J's proposal.

L&J offer evidence that, at a visual level, objects and their relations are captured using "spatial representations", chiefly expressed by adpositions. Size, orientation, curvature and other physical properties all conspire for an object to be recognized as more than a sum of its parts: a 'whole' entity, or what the object is. Whole objects or '*what*s' can also be related one to another: If we have two objects, one will be conceived as a landmark object (or *ground*), while the other will be the 'located' entity (or *figure*; Talmy 1978, 2000).

They also argue that the rich and variegated layers of visual-cognitive information are processed and then clustered together and associated with 'conceptual labels' (or just 'concepts') and hierarchically organized within the

Conceptual System (CS, Jackendoff 1983, 1990, 1991, 2002), the interface between non-linguistic modules and (linguistic) domain of semantics. This proposal and further extensions assumes that *nouns* are the main category representing objects in language, whereas *adpositions* represent spatial representations/relations (e.g., van der Zee 2000). In line with other literature, L&J propose that spatial expressions mostly involve 'count' nouns, which can be seen as labels for objects with a given 'shape' (e.g., 'cylinder' or the fictional 'dax': Carey 1992, 1994, 2001, Soja *et al.* 1992, Bloom 2000, Carey & Xu 2001). Adpositions, on the other hand, are argued to express core geometrical properties such as overlapping, distance and orientation (e.g., *in*, *in front of*; Landau & Stecker 1990, Landau *et al.* 1992).

Recent inter-disciplinary research has shown that the picture is somewhat more complex. A rich body of evidence has been accumulated suggesting that adpositions can also convey information which is not necessarily geometric in nature. Look at the examples:

(6) The book is on the table.

(7) Mario is beside the table.

(8) #The table is beside Mario.

(9) Mario is taking the moka machine to the kitchen.

If a book is "on" the table (as conveyed by (6)), the table will also act as a mechanical support to the book, that is, it will prevent the book from falling. We can say that Mario is "beside" the table (as in (7)), but saying that the table is beside Mario will be pragmatically odd (as in (8)):[3] Figures tend to be animate entities (or at least conceived as such), whereas grounds tend to be inanimate entities.

These mechanical properties can also be seen as extra-linguistic or 'spatial' properties associated to nouns. Informally, if a count noun such as *book* is associated to an object with definite shape, it can (and should) be involved in causal physic relations (e.g., support, or containment); cf. Kim & Spelke (1992, 1999), Spelke & van der Walle (1993), Spelke *et al.* (1994), van der Walle & Spelke (1996), Spelke & Hespos (2001), Smith *et al.* (2002), Shutts & Spelke (2004).

Dynamic contexts offer similar evidence for the relevance of extra-geometric information to be relevant. For instance, in a scenario corresponding to (9), we will understand that the Moka machine[4] brought to the kitchen by Mario will reach the kitchen because of Mario's action (Ullman 1979, 1996, von Hofsten *et al.* 1998, 2000, Scholl 2001, 2007). We will also take for granted that the machine's handle and beak will reach the kitchen as well, as parts of the machine, unless some problem arises in the meanwhile. If Mario trips and the Moka machine falls mid-way to the kitchen, breaking in many pieces, we may not be able to recognize the Moka machine as such (Keil 1989, Smith *et al.* 1996, Landau *et al.* 1998). Spatial relations, and thus adpositions that express these relations, can implicitly capture the (potential) causal relations or affordances between dif-

[3] Examples (7) and (8) and related discussion are based on an issue correctly pointed out by an anonymous reviewer, whom I thank.

[4] The traditional Italian machine for espresso coffee.

ferent objects (e.g., Landau 1994, 2002. Munnich & Landau 2003).

For these reasons, Coventry & Garrod (2004) propose their FGF framework, according to which mechanical, geometrical and affordance-oriented properties form the *mental model* or *schema* (in the sense of Johnson-Laird 1983) of adpositions that we store in *long-term memory*. This model can be seen as the 'complete' representation of an adposition's meaning, which can then only partially correspond to its actual instantiation in an extra-linguistic context (see also Herskovits 1986).

According to this theory, speakers can then judge a sentence including a spatial adposition more or less appropriate or felicitous, depending on whether the adposition's content is fully or partially instantiated in an extra linguistic scenario (e.g., van der Zee & Slack 2003, Coventry & Garrod 2004, 2005, Carlson & van der Zee 2005, Coventry *et al.* 2009, Mix *et al.* 2010). Two examples are the following:

(10) The painting is on the wall.

(11) The painting is in the wall.

A sentence such as (10) can be considered more appropriate than (11) when used in an extra-linguistic context in which a certain painting is just hanging on the wall, but less appropriate when the painting is literally encased in the wall's structure.

Other theories take a perspective which is either close to L&J or FGF. *Vector Grammar Theory* (O'Keefe 1996, 2003) treats English adpositions as conveying information about vector fields, the graded sequence of vectors representing the minimal 'path' from ground to figure, and thus conveying purely geometric information. Another theory which is based on similar assumptions is the *Attentional Vector Sum* model (AVS; Regier & Carlson 2001, Regier & Zheng 2003, Carlson *et al.* 2003, 2006, Regier *et al.* 2005). In this theory, 'vectors' represent features of objects that can attract the speaker's attention once he interprets a spatial sentence, and can thus include mechanical and functional aspects as well as environmental ('reference frames') information.

These theories thus predict that a sentence such as (12),

(12) The lamp is above the chair.

is interpreted as a 'set of instructions' that informs us about where to look at, in a visual scenario, but they differ with respect to these instructions being purely geometrical or not. Furthermore, AVS predicts that *above* will be considered more appropriate if used in an extra-linguistic context in which the lamp is above the chair also with respect to three possible systems of orientation or reference frames, for example, if the lamp is above the chair with respect to some environmental landmark such as the floor (*absolute* reference frame), with respect to the chair's top side (*intrinsic* reference frame), and with respect to the speaker's orientation (*relative* reference frame); see e.g. Carlson-Radvansky & Irwin (1994), Carlson (1999).

Although the insights from these theories are quite enlightening and consistent with various approaches to vision, their approach to language is inherently a 'blurry' one, as each of these theories says virtually nothing about the specific contribution of nouns and adpositions. Since these theories tend to reduce language to general cognition, this is not surprising. Aside from this problem, no theory really attempts to analyze 'dynamic' spatial expressions. The same holds for L&J and FGF: Examples such as (4) and adpositions such as *to* are still a mystery, with respect to the vision–language interface. Nevertheless, both sides of the debate offer at least two important points regarding the nature of spatial vision and spatial language.

These aspects form the answer I shall propose to the first research question:

A–1: *Previous literature offers a clear mapping between vision and language (L&J), and evidence that spatial vision and language express possible relations between entities (FGF).*

Because of these previous proposals I shall assume, based on the literature on the topic, that spatial vision and spatial language are not just about geometrical relations, and thus suggest that both modules can express the same 'amount' of spatial information, although in (quite) different formats. I shall also assume that there is one precise, although flexible, correspondence between units of vision and units of language. Visual objects find their way in language as nouns, and spatial relations as adpositions, at least for English cases I shall discuss here.[5] In the next section, I shall offer a justification to these assumptions and propose a richer theory of spatial vision and language.

3. The Nature of Spatial Vision and Language, and a Formal Analysis

In this section I shall offer an analysis of 'static' and 'dynamic' vision (sections 3.1 and 3.3), and a logic of vision of these theories (sections 3.2 and 3.4); I shall then analyze (specific aspects of) spatial language via DRT (section 3.5).

3.1. Classical and Modern Varieties of Object Recognition

In highly schematic terms, we can say that spatial information is processed via visual perception, for most human beings. Light 'bounces' off an object and the surviving wave-length is processed by the eyes. This information is then transmitted to the optic nerve, to be further processed in various parts of the brain, like the primary and secondary visual cortex. Once the perceptual inputs are processed, their corresponding (internal) representations become the basic chunks or atoms of information processed by higher cognitive functions, such as vision and

[5] A specific language may lack a term for a certain visual object, so the correspondence between visual objects and nouns on the one hand, and spatial relations and adpositions on the other hand, may be subject to subtle cross-linguistic variation. Informally, if a language has a term for a certain visual object, this term will be a noun, syntax-wise: The same holds for spatial relations. I thank an anonymous reviewer for bringing my attention to this point.

memory.

One of the earliest schools of research that attempted to investigate the nature and properties of these units of information was the *Gestalt* school of psychology. This school assumed that our unconscious processes of visual recognition allow us to individuate objects from the background via the following four principles: *invariance* ('sameness' of an object), *emergence* (parts making up a whole), *reification* (interpolation of extra information), and *multi-stability* (multiple 'good' images of an object). These principles converge into underlying principle of *Prägnanz* or *conciseness*, our ability to form discrete visual units from different, and perhaps contradictory, 'streams' of perceptual information. This process may not necessarily be 'veridical' in nature: If we look at a car in motion and we do not notice its radio antenna, we may consider the two objects as one, as long as there is no visual cue that they are indeed distinct objects (e.g., the antenna breaks and flies away).

The Gestalt school's thrust in the study of invariant properties lost momentum after the end of World War II, until Gibson (1966) re-introduced the study of vision as a process of 'information-processing' (and integration), which sparked the interest of various researchers,[6] including David Marr and his model of vision which had an ever-lasting influence in vision sciences and in some linguistic literature (e.g. van der Does & van Lambalgen 2000).

Marr's initial research started from the physiological bases of vision (collected in Vaina 1990). His interest slowly shifted from the neurological and perceptual facts to cognitive aspects of visual processes, which culminated in Marr (1982). The core assumption in Marr's theory is that vision can be best understood and represented as a computational, algebraic model of information processing. It is a bottom-up and cognitively impenetrable process, since it is mostly realized without the intervention of conscious effort.

Marr proposed that any model, and thus any mental process or structure it represents, should be defined at three levels of understanding: *computational* ("why" of a model), *algorithmic* (the "how" of a model), and *implementational* (the "what" of a model). Marr proposed that our vision developed with a perhaps very abstract computational nature, that of 'grouping' any type of visual information (geometric and not) into implementable units, which can be retrieved and stored in memory. Regardless of its purposes, Marr proposed that the computational system of human vision is assumed to have three intermediate levels of representation, or 'sketches'.

At the *Primal Sketch* level, boundaries ('zero crossings') and edges are computed, so that the continuous stream of perception is partitioned into discrete units of attention, or 'receptive fields'. Photo-receptive cells detect the change of light in the receptive fields, and split it in two parts: an 'on-center' and an 'off-center'. In 'on-center' cells, the cell will fire when the center is exposed to light, and will not fire when the surround is so exposed. In 'off-center' cells, the opposite happens. When both types of cells fire at the same time, they are able to

[6] J.J. Gibson would come to reject his stance in favor of an 'ecological' or 'externalist' approach, in Gibson (1979). More information about perceptual and historical aspects can be found in Scholl (2001, 2007), Bruce *et al.* (2004), and Farah (2004), *inter alia*.

represent an entity like an edge, its adjacent 'empty' space and the boundary between the two partitions. The change of polarity between these two partitions is defined as a *zero-crossing*. A zero-crossing represents change in terms of opposed polarities: if an edge is marked as +1 in value, then the adjacent 'empty' part will have value –1, and a border will be represented as 0, or as a 'boundary'.

At the *2½-D sketch* level, these elements are integrated in the computation of surfaces and their distance from the observer. For instance, a triangle represents three lines whose edges coincide in a certain order, forming a connected contour, the triangle itself. Other information, such as depth or orientation, is computed via the integration of information about, respectively, the distance of the single surfaces from the observer (hence, an *egocentric* perspective), and integrated in a mean value, the normal 'vector' from those surfaces. Missing information can here be interpolated: If part of the triangle's side is occluded, we may just 'infer' it from the orientation of the visible sides.

At the *3-D model* level, the recognized parts and portions are integrated into one coherent whole. At this level, vision becomes an object-centered (or *allo-centric*) process, which allows for shape recognition to be viewpoint-invariant. The computation of a full 3-D model (object recognition) is crucially based on how the computation evolves from the 2½-D sketch to its final level. If the various 2½-D sketches can be integrated into a coherent unit, and this computed unit matches with a corresponding unit in memory, then the process of 'object' recognition is successful (see also Marr & Nishihara 1978).

Marr's model, given its algebraic nature, can be informally stated as a model in which basic information units or indexes can represent single parts of an object: *a* and *b* can stand for head and torso of a human figure, represented as the index *c*. If the unification or merging[7] of the two more 'basic' information units *a* and *b* into a single unit is identified with a whole, then object recognition occurs. Simply put, from head and torso (and other parts) we obtain a human figure, a process that can be represented as $(a+b)=c$, c standing for the human figure index.

This quite informal exposition should already made clear that two basic principles can be identified as being part of spatial vision. One is the need to 'chunk' the perceptual stream into discrete, computational units; and the other possibility to 'merge' and identify these units in a rather abstract way, which allows us to establish *part-of* relations, according to Marr, among different information units.

After Marr's seminal work, theories of object recognition roughly distributed between a more representational and a more derivational stance. While representational theories stress relations between different objects and parts (or, rather, representations thereof), derivational theories stress the processes by which these representations come into being. I will start from the representational stance, introducing *Recognition By Components* theory (henceforth RBC; Biederman 1987, Hummel & Biederman 1992), probably the most influential theory for the representational stance.

---

[7] Here I use the term 'merge' in a pre-theoretic way, but I will offer a more precise definition in section 3.3.

RBC offers an approach which is substantially similar to Marr's original proposal, although it is postulated that object recognition occurs via 7 sketches of representation, rather than 3. One important difference is that, after the first two sketches are computed, each (part of an) object is conceptualized as a *geon* (*generalized ion*; Biederman 1987), a primitive shape or visual 'ur-element'.[8] The combination of various geons allows to define complex forms: For instance, an ice-cream can be idealized as a semi-sphere connected to a cone, consequently capturing complex relations between the parts they represent. Whenever an object is successfully recognized, it can be and stored in memory as a distinct entity (Hummel & Stankiewicz 1996, 1998, Stankiewicz & Hummel 1996).

An important aspect of RBC is that it addresses how different information units are combined together over the time of a computation, a phenomenon defined as dynamic binding. Informally, if we recognize a sphere shape *a* and a cone shape *b* at a(n interval) time *t* in the computation, their integration as integrated units *a+b* will occur at a time *t+1*. In this perspective, object recognition can be seen as a dynamic process of binding different units of information together, so that 'new' objects emerge from this process: By dynamically binding edges and lines together in a coherent representation we have surfaces, and by dynamically binding surfaces together we have three-dimensional objects, at an interval *t+n*.

An alternative view to this representational approach may be exemplified by the derivational model *H-MAX* (short for 'Hierarchical MAXimization' of input) of Tomaso Poggio and associates (Poggio & Edelman 1990, Riesenhuber & Poggio 1999a, 1999b, 2000, 2002, Serre *et al*. 2005). In this model, objects can be any parts of which we receive visual input, via their luminosity, and of which we compute possible visual candidates (e.g., different possible representations of the same dog). No intermediate levels of representation are however assumed to exist, since the flow of information is constrained via a pair of simple principles, *SUM* and *MAX*, which are in turn defined over vectors as sequences of minimal parts and boundaries of an object.

An example is the following. Suppose that we look at our pet Fido, starting from his tail. At this initial step, our visual system first computes parts and boundaries, such as the tail's tip, which can be badly lighted or 'stilted', if we are observing it by an odd angle. From this 'vector', we access other possible memorized images of Fido's tail and combine them with other visual features (vectors) we recognize about Fido. In case the image is somehow poor, we may compare it as a 'noisier' version of Fido's tail.

All these vectors are then summed together in the sum vector, the averaged sum of the vectors corresponding to the various visual inputs. If this sum exists, then a 'standard' (or allocentric) view will be defined, which corresponds to the final step of the process of object recognition. In keeping track of these different views, 'feature clusters', edges of a surface or other easily observable points play

---

8 Geons are not exactly primitives *per se*, but represent the (finite) set of combinations (36 in total) of 5 binary or multi-valued properties that combine together to define a shape. These five properties are: *curvedness* (if a component is curved or not), *symmetry*, *axis* (specifically, the number of axes), *size*, and *edge type* (if the edges define an abrupt or smooth 'change of direction').

a vital role.

In more formal terms, the *SUM* takes two visual objects and unites them together into a new visual object: If a and b are Fido's head and torso, then $a+b=c$ is Fido's body. The *MAX* operation minimally differs from the *SUM* operation in two subtle ways. First, it may sum together two visual objects and obtain one of the two objects as the result, i.e. $a+b=b$. This is possible when one object 'includes' the other, i.e. when one visual object contains all the features of another object; hence, their union will be the 'strongest' object. Second, it may average visual objects representing the same entity, i.e. it may sum objects which have common features. In formal terms, this can be then represented as $(a+b)+(b+c)=a+b+c$, a novel visual object (the 'average' image) obtained out of previous objects. These processes are dynamic, so if two visual objects are *SUMmed* (*MAXed*) at a time t, the result will hold at a time $t+1$.

While these two theories show a substantial convergence in their treatment of object recognition, their assumptions about the nature of 'objects' is quite different. Representational theories consider an 'object' as the end result of a visual computation, while derivational theories consider an 'object' as any unit that is manipulated by a computation. This difference may appear purely theoretic, but it has its own relevance once we take in consideration how this information is mapped onto linguistic units. Consider, for instance, the following examples:

(13) The book is on the tip of the left edge of the blue table.

(14) The book is on the table.

In (13), the spatial relation is defined over a book and a rather specific part of a blue table, the tip of its left edge, whereas such level of detail is left implicit in (14). Note that this relation also informs us that the book is supported by one part of the table (the tip of the left edge), which in turn may be seen as not so ideal for supporting books (tips are intuitively worse 'supports' than centers).

For the time being, though, I shall leave aside adpositions and spatial relations, and concentrate on objects and nouns. In both sentences, any object or part thereof ('edge', 'tip') finds its linguistic realization as a noun: If there is a difference between different layers of visual representation, this difference disappears at a linguistic level, since both visual objects are represented in language as nouns. Consequently, a theory of object recognition that makes no difference between parts and whole objects, such as *H-MAX*, offers an easy counterpart to these simple linguistic facts, while other theories are less suitable for my goal of offering a theory of the vision–language interface. I shall base my formal proposal on vision by offering a logical treatment of *H-MAX*, in the next section.

3.2. A Logic of Vision, Part I: Static Vision

The core aspects shared by the models of static vision (object recognition) we have seen in the previous section are the following. First, vision involves the explicit, internal representation of perceptual stimuli in terms of discrete information units, or visual objects (of any size and shape, so to speak). Second, these units are combined together via one underlying principle, which we can temporarily label as 'sum'. Third, the result of this process defines more complex objects, but also relations between these objects, which can be seen as instances of the *part-of* relation. These three aspects can be easily represented in one (preliminary) unified logic of vision, which I shall define as follows, and which I shall expand in more detail in section 3.4.

First, I shall assume that vision includes a set of *visual objects*, the (countably infinite) set $V=\{a,b,c,...,z\}$. Each of these objects represents a minimal information unit, an output which is activated (instantiated) when some perceptual input exceeds a threshold level. Hence, each information unit in a computation represents an instance of *transduction*, since it represents the (automatic) conversion from one type of (input) information to another type of (output) information (Pylyshyn 1984, Reiss 2007). I shall assume that each object can be represented as a singleton set, via 'Quine's innovation': Hence, a is shorthand for $\{a\}$; consequently, our operations will be defined over sets (cf. Schwarzschild 1996: appendix).

Second, I shall assume that one syntactic operation can be defined over these units, the sum operation '+', an operation that I will call *merge*. An example of merge is $a+b=c$, which reads: "c is the merge of a and b". It is a *binary* operation which is also *associative*, *commutative*, and *idempotent*. Associativity means that the following holds: $a+(b+c)=(a+b)+c$. In words, and using again the example of Fido, Fido's head with Fido's body (torso and legs) correspond to the same object as Fido's upper body and legs: Fido. Commutativity means that the following holds: $a+b=b+a$. In words, Fido's head and body form Fido, much like Fido's body and head. Idempotence means that the following holds: $b+b=b$. Fido's head and Fido's head give us Fido's head, that is, we can repeat information. Since our objects are singleton sets, this operation is basically equivalent to *set union*. The intuition behind the *merge* operation is that it takes two 'old' distinct objects and creates a 'new' object as a result, in a sense distinct from the basic sum of original parts. For instance, our Fido can be conceived as the new visual object that is obtained when the visual objects corresponding to Fido's body and Fido's head are merged together into an integrated representation, Fido as a 'whole' entity.

Third, I shall assume that one semantic relation can be defined between objects, the *part-of* relation, represented as '≤'. An example of the *part-of* relation is $a \leq b$, which reads: "a is part of b". Since I am using Quine's innovation, the *part-of* relation is roughly equivalent to set membership.[9] This relation is also binary, and it is *reflexive*, *transitive* and *antisymmetric*. It is reflexive, since the following holds: $a \leq a$. It is transitive, because the following holds: if $a \leq b$ and $b \leq c$, then $a \leq c$. It is antisymmetric, because the following holds: if $a \leq b$ and $b \leq a$, then $a=b$. In words, each part of Fido's is part of itself (reflexivity); if Fido's leg is part of Fido's body and Fido's body is part of Fido, then Fido's leg is part of Fido (transitivity); if Fido's body parts are part of Fido, and Fido consists of Fido's body parts, then

9 The subtle but very important differences between the notion of 'set membership' and the *part-of* relation are not important for our discussion. However, the interested reader is deferred to e.g. Link (1983, 1998), Landman (1991: chap. 1), Schwarzschild (1996: chap. 1) for discussion.

they are recognized as the same entity (antisymmetry). The intuition behind the *part-of* relation is that it establishes a relation between 'old' objects and a 'new' object as a result of the *merge* operation. For instance, if Fido is the result of merging Fido's legs and Fido's body into a 'new' object, then Fido's legs will be part of Fido. If we recognize Fido, then we will also recognize Fido's legs as well as other parts that make up Fido, as a consequence of the relation between parts and whole.

The resulting model of (object) vision emerging from these basic definitions is the triple $S=<V,+,\leq>$, a simplified variant of a structure known as join lattice, a type of full Boolean algebra (e.g. Keenan & Faltz 1985: chap. 1, Landman 1991: chap. 2, Grätzer 1978: chap. 1–2). A join lattice can be seen as a set with at least one binary operation of composition and one relation defined over its elements, which also has the following property: if $a\leq b$, then $a\cap b=a$ and $a\cup b=b$. In words, if *a* is part of *b*, then the intersection of *a* and *b* is *a*, while the union of *a* and *b* is *b*. Informally, if the merge of two objects creates a novel object, the part of relation establishes that this novel object includes the old objects as its (proper) parts. Because of these properties, this type of Boolean algebra is a complete structure, i.e. it will have one maximal object including every other object (i.e. *V*) and one minimal object which is included in every other object, which we will call '*0*', and which represents any instance in which we 'fail' to recognize objects.[10]

Since we mostly operate on individuals, i.e. singleton sets via *merge* and the *part-of* relation, the logic of vision I define here is substantially a first order logic. Since this logic allows us to define an algebraic model of objects and their interpretation and relations, it is a model-theoretic approach to vision. Anticipating matters a bit, the discussion of the vision–language interface will coincide with the discussion on how this model and the model defined by language are related.

These logical/algebraic properties represent the following facts: The visual 'integration' of Fido's leg and Fido gives us Fido, i.e. Fido's leg is 'recognized' as part of Fido's whole image (union). If from Fido's whole image we focus on Fido's leg, then the other parts will be ignored (intersection). This latter interpretation of 'attention as intersection' can be found in RBC and Ullman (1996), and is based on one simple intuition: If *merge* represents object recognition (the union of different visual) inputs, then its complementary operation represents the process by which we focus on a single visual object out of an array of objects, i.e. attention. Furthermore, the sum of objects forms the full 'library' of our model of vision (the maximal object *V*), and there can be cases in which we cannot recognize any object whatsoever, for instance when we fail to focus our attention on something (the empty object).

This brief and semi-formal *excursus* suffices for our discussion of object recognition. The important aspect is that we can now define a tight relation between the syntax and semantics of our logic of vision: For each instance of the

---

10 Note that the operation *MAX* can be now reconstructed as a special instance of *SUM* (i.e. our merge). I shall leave to the reader the simple proof of this fact. Also, note that given our definition of the sum operation, visual objects can be either *atomic*, i.e. they only include themselves as proper parts (e.g., *{a}*), or non-atomic, they may have other objects as their proper parts (*plural/sum* objects: e.g., *{a,b}*, including *{a}* as its part; see e.g. Link 1983, 1998 and Schwarzschild 1996). The import of this subtle distinction is not crucial, in this paper.

merge operation, the result will define another visual object and a *part-of* relation between this object and its constituent parts. Informally, we are able to recognize the legs of Fido as Fido's, because we first integrate Fido's legs with other Fido's body parts into Fido's whole image, and then retrieve this relation between legs and Fido.

The merging of visual objects does not occur in a temporal void, as we have seen, but is dynamically realized over discrete intervals of time. In RBC, this is represented via *dynamic binding*, i.e. the explicit representation of derivations as they occur over time. Before defining dynamic binding, I shall define the structure of the Index Set that represents intervals of time. This structure is the duple $I=\langle t,+\rangle$, a set of intervals of time with an operation of *addition*. Although I represent this operation via '+', it is a slightly different operation than *merge*, since it is only associative but not commutative nor idempotent. Intuitively, from a starting interval t we can 'move forward' to other intervals, e.g., $t+1$, $t+2$ and so on, via the simple iteration of this 'asymmetric' *merge*.

The corresponding type of structure is a simpler algebra, a *total order*, i.e. a structure in which each element is a distinct object. Intuitively, this structure represents the directed flow of the logical processes underpinning visual computations, the 'arrow of time' that tells us how visual objects are integrated together, but which cannot 'remember' any relations between the objects manipulated in these operations.

The explicit integration of this structure with vision is the duple $S_d=\langle I,S\rangle$, the 'dynamic' logic of vision and object recognition. Its dynamic nature stems from the ability to represent visual computations as they occur over derivational times, in a simple format similar to standard proof-theoretic (i.e. syntactic) component of various logical systems (see e.g. Landman 1991 for discussion). One example is the following:

(15)	t.	a	(visual object instantiation, e.g. Fido's head)
	$t+1$.	b	(visual object instantiation, e.g. Fido's body)
	$t+2$.	$a+b$	(*merge* introduction)
	$t+3$.	$(a+b)=c$	(Fido as 'sum' of Fido's parts)
	$t+4$.	$a\leq c$	(*part-of* introduction, Fido's head as part of Fido)

This derivation roughly captures how the process of recognizing Fido may occur a dynamic (and bottom-up) way, modeling the actual processes described in the reviewed theories. The various objects are first recognized ('instantiated' in the derivational space) one by one and then merged via the introduction of this operation. Once this process is over, we can also access the relation between Fido's head and Fido's whole image, since we can establish that one is part of another.

This simple example of a derivation in our logic of vision may not capture all the aspects involved in visual computations and, to an extent, it is quite idealized: For instance, an individual may consciously assume (and thus exert a *top-down* choice) that he is seeing Fido's body, since he can partially recognize it as a visual entity connected to Fido's head. In this and other examples, I shall leave these matters aside, as they are not crucial, for our discussion. This example, however, introduces one important advantage of my theory over the theories I reviewed so far: it makes fully explicit the structural relations between the various components of the object recognition process, including its unfolding over time. This logic of vision is still a preliminary proposal, since for one thing, it does not allow us to make a distinction between objects (individual constants such as e.g., *a*) and the properties they instantiate (e.g., constant functions such as **dog'**). It also cannot represent spatial representations, and thus the visual content of adpositions, but this is a void that will be filled in section 3.4, along with a theory of visual properties. However, it already allows us to give a compact definition on how we see things in the world, at least with respect to static objects.

Now we can explicitly represent (visual) objects in a very preliminary *logical space*, and we can also define how these objects are mapped onto their corresponding linguistic labels, nouns. I shall assume, differently from previous proposals such as L&J, that this mapping is an *isomorphism*, a one-to-one correspondence between objects of different types (i.e. visual objects to noun labels). The reasons for this assumption are the following. The discussion of examples (13) and (14), and the intuition that each visual object may (potentially) have a corresponding 'noun' label, has one important theoretical consequence. If we define a function mapping visual objects to nouns, then this function will be *injective*, it will find at least a label ***n'*** for each visual object *v*: A noun like *table*, for instance, stands for the corresponding visual object, a table. Furthermore, it is possible that several visual objects can correspond to one linguistic label: A noun such as 'table' also stands for the sum of legs, surface, edges, and other visual objects making up a table. Hence, this mapping function will be *surjective* as well.

A function which is injective and surjective is a *bijective* function, hence a function that defines an isomorphism. More formally, for each visual object *v*, for each noun label ***n'***, there will be a function *f* such that : $f(v)=\mathbf{n'}$**.** Since this function is surjective, the following holds: given $a+b+c=v$ then $f(a+b+c)=\mathbf{n'}$. In words, we have the 'lexical' identity $\mathbf{edge'}+\mathbf{legs'}=\mathbf{table'}$, which can be also indirectly represented as $f(a+b)=f(a)+f(b)$, with $f(a)=\mathbf{edge'}$, $f(b)=\mathbf{legs'}$ and $f(a+b)=\mathbf{table'}$. Furthermore, this isomorphism preserves relations, so if one object is part of another, one corresponding noun will be lexically related to another. We have $f(a)\leq f(b)$, which in words says that *edge* is (lexically) related to *table*.

This isomorphism can be interpreted as follows. Our logic of vision is a partial, yet very fine-grained model of object recognition, with a simple yet rich *hierarchical structure*, defined by the *part-of* relations that can be established between the objects in this domain. The function *f* tells us that such structure can also be connected with other structures, provided that they are governed by the same (logical) principles. Informally, it allows us to potentially define a correspondence between nouns in language and visual objects in vision, on a one-to-one basis. Although a language may lack a specific lexical item for each visual object, it is at least possible to define such a tight correspondence between nouns on the one hand, and visual objects on the other hand.

This function can be thus thought as representing a top-down, conscious (and optional) process, which occurs when we consciously match visual information against linguistic information. It allows to define a correspondence between simple and complex visual objects and the nouns that represent these objects at a linguistic level, e.g., to establish that a noun such as table can indeed refer to[11] a visual object we may observe, and which is made of four legs, a surface and other relevant parts. With this notion in place, then, we have introduced enough 'machinery' to handle the static side of vision and its logic; we need to focus on the neglected dynamic side, and propose a full logic of vision, by which we can also analyze spatial representations/relations. I shall do so in the next two sections.

3.3. Theories of Dynamic Vision

In the discussion in the two previous sections, I have introduced a view of spatial vision in which the ability to explicitly represent objects and their relations plays a crucial part in 'static' scenarios, i.e. cases in which we 'find' objects which are not changing position over time. One aspect missing from this discussion is how we establish relations between objects, especially when they change position over time — how dynamic spatial vision comes about.

A preliminary step to answer these questions is to define how we can keep track of objects over time. For this purpose, I shall review a theory about dynamic object tracking: *Multiple Object Tracking* (MOT), introduced in Pylyshyn (1989) and developed in a number of successive works (e.g. Pylyshyn 1994, 2001, 2003, 2004, 2006, and Pylyshyn & Annan 2006; see Kahneman *et al.* 1992 for the roughly equivalent *Object File Theory*).

MOT offers a theory about object recognition in dynamic scenarios by analyzing how we are able to individuate and form mental representations of objects in the way they instantiate some properties (e.g., being yellow in color), and by how we maintain or change these representations over time and the unfolding of events. MOT is probably best presented via a preliminary example. Imagine that we look at the panorama: We detect trees, clouds, buildings, and so on. If we focus our attention on a flying black swan, we can do so because we are first able to detect a mysterious object (call it '*x*'), which instantiates the properties "swan", "black", and "flying", among others.

With some imagination, we can assume that "swan" is the primitive and most basic property which allows us to recognize the mysterious entity as such, the equivalent of an imaginary finger stretching from our eyes to the object itself. Such a finger allows us to define the mysterious object in terms of what property it instantiates, and it is thus defined as *Finger of INSTantiation*, or *FINST*. The very act of this process is usually defined as *FINSTing* in the literature and, since it can be defined for any entity that can be so individuated, it makes no distinction between types of objects: Everything which can be *FINSTed* is an object, simply enough.

It is useful to illustrate MOT's notation for the basic process of FINSTing, as well as the addition of further features. I will follow Pylyshyn's (1989) notation, for ease of exposition. Aside from the basic process of FINSTing, we can imagine

---

11 The notion of 'reference' I use here is not equivalent to the one commonly employed. A standard assumption is that reference is the relation between a term and the 'real world' object that corresponds to a given term. Here and for the rest of the paper, I shall assume that linguistic terms can refer to extra-linguistic but internal information, such as visual objects.

a situation in which the black swan is flying above a cloud. The process of FINSTization is illustrated in (16), while (17) illustrates the more complex 'above' case:

(16) a. *FINST[x],[swan]=(x:swan)*
b. *FINST[x:swan],[x:black]=(x:swan,x:black)*

(17) *ABOVE(x:swan, x:black, x: f lying, y:cloud)*

In (16a), a basic property like "swan" is mapped onto a visual object, acting as the FINST that tracks the visual object. In (16b), the combination of two properties acting as FINSTs creates a new, more complex FINST, which identifies the visual object x as a black swan. In the case of (17), we can observe that such 'internal fingers' can also define relations between simpler 'fingers', hence expressing a relation between different instances of the same underlying process.

This relation is, in turn, a description or property of an event of motion, in which the swan is the moving figure, while the cloud is the contingent ground. Further information can be stacked up via dynamic binding: Informally, each individuating property for x can be in a temporally incremental fashion (e.g., "black" at time t, "flying" at time $t+1$), which in turn is realized via the iterated application of the FINST operation.

One problem emerging from the presentation of MOT is that this theory cannot easily be used to analyze how the temporal relations between properties can be defined and represented in their own right. While "black" may be instantiated after "swan", we cannot explicitly represent that the corresponding 'fingers' can be taken as entities in their own right, the events during which these properties are instantiated and combined together, or defined in terms of their order of occurrence.

One theory that aims to fill this conceptual void is *Event Segmentation Theory* (henceforth: EST), a theory of events and psychological events first outlined in Zacks & Tversky (2001) and Zacks *et al.* (2001). In this theory, an original philosophical intuition by Quine (1960) and further developed in Davidson (1967) acts as the basic insight and ontological assumption: that our understanding of the world includes not only objects, but also the events in which these objects are involved.

At one level of comprehension, our mind represents objects as "things that are in the world", such as birds and apples and cups. Once we add a temporal level of comprehension, and thus we observe how things change or preserve their own identity through time, we also keep track of what causes may change the properties of an object. The focus of EST is on events, which are treated as 'pegs', basic computational units or 'slots' on which we stack up information, and which stand for relations and order among relations in which objects are involved, as they unfold over time (Speer *et al.* 2003, Zacks 2004, Zacks *et al.* 2007, Zacks & Swallow 2007, Reynolds *et al.* 2007, Tversky *et al.* 2008).

EST assumes that, at a minimum, we can observe objects at two levels. One basic level is that of their structure and how it is realized in space (*partonomy*) and one of an object and its relation to an abstract class (e.g., a chair as part of furniture: *taxonomy*). Once we take in consideration a temporal dimension, in which objects can have different properties in different intervals of time, we will have 'dynamic' objects or *events*. Events are conceived as discrete information units derived (i.e. *transduced*) from perceptual information, i.e. the 'indexes' attributed to the combination of a (rather abstract) visual property and the object that instantiates it.

For instance, if someone throws a cup on the floor, then the cup will likely be shattered into pieces because of this action. The temporary relation between an individual and the cup will bring about a new state of affairs in which the cup will be a new property of some sort, that of being shattered. At the same time, we represent this change via the *temporal* and *causal* relation between the two state of affairs, one involving an event of someone shattering the cup, and another in which the cup will be shattered, which is separated by a boundary event, an interval of time in which neither the cup is shattered nor it is still intact, and in which we will need to 'update' our model of events. Events can also be combined together: If someone is stacking pillows, each single pillow-stacking event can be combined into a 'bigger' pillow-stacking event, and possibly 'merged' with other events, forming a more complex event such as 'pillow-ordering'.

Such complex sequences of events can be seen as event models or schemata, structures of events and their causal/temporal connections, as they are represented in *short-term memory* (models) or stored in *long-term memory* (schemata in 'semantic memory': see e.g. Tulving 1972, 1983, 2000a, 2000b, 2002 and references therein for an introduction). Events can be dynamically bound: The "throwing" event occurs at a time *t+1*, a boundary event is formed at a time *t+2* and the "shattering" event occurs at a time *t+3*, then there will be a causal, as well as temporal relation between these events.

Both MOT and EST are theories that offer a detailed picture of how dynamic vision can occur, defining in detail the mechanisms by which we track objects in motion, and the complex spatial representations that arise from this process, or events. One alternative view to these approaches that offers some further important insights on spatial representations is the *Hippocampus as a Cognitive Map* theory (HCM) of O'Keefe & Nadel (1978). HCM started as a study of rats' navigational system, the way they represent objects and their places in the environment, and how this information is memorized and accessed or updated at later stages. According to this theory, humans (and rats) build up a complex spatial representation of the environment via two parallel systems: the *place* and the *misplace* system. The place system records information about objects' position in the environment, and 'checks' whether this information is correct when visual information is processed. If an object has changed position, then the misplace system records the change of position and updates the object's new position accordingly.

This model has been further extended over the years. O'Keefe (1983, 1990, 1991) and Burgess & O'Keefe (1996, 2003) show that information about objects and their relations is processed, in 'real time', by the navigational system. This system computes the location of a figure in terms of polar angle Φ and distance d from the ground as the relation $\theta(\Phi,d)$, computed via θ*-rhythm* signals, which mostly originate in the Hippocampus.

The result of these computations can be modeled as a vector, a sequence of cells (*Boundary Vector Cells*) that fire when an observer visually tracks relevant entities in an environment, and can also allow to compute geometrical properties of objects. Hence, the place and misplace systems build up complex spatial representations over time, or Cognitive Maps (O'Keefe & Burgess 1999, Burgess *et al.* 2002, and Burgess 2006a, 2006b; see Arsenijević 2008 for a linguistic proposal).

These theories give us a basic insight on the nature of dynamic spatial vision. When we keep track of objects in motion, we do via the properties that objects may have over time, whether they are geometrical, functional or 'functional', insofar as they allow us to track objects in space. At the same time, we also keep track of the relations between these properties and their order of causal/temporal occurrence: Spatial representations have an inherent temporal dimension, which represents the structural relations between the events making up these representations.

Adpositions, as the chief part of speech expressing these relations, must also have such an abstract nature. Look at the examples:

(18) Mario has fallen onto the floor.

(19) Mario has gone into the room.

(20) Mario is sitting near the patio.

A scenario which is more or less depicted by (18) and (19) is one in which Mario is on the floor and in the room, respectively, as a consequence of a particular event, one of falling and one of going. A scenario depicted by (20) is one in which Mario is involved in an event of sitting, which occurs at some distance from the patio. He may be involved in other events, although these events are in a sense 'backgrounded', in the sentence.

In all three cases, the spatial relation holding between Mario and different grounds holds at some moment of time because of some previous event, and involves more than just geometrical information. If we conceive Mario and the floor as inherently visual objects, then the adposition *onto* will capture not only that these two objects are currently related one another via a 'support' relation, but also that such relation has come into being because of a previous falling event. Since adpositions seem to express the 'logical' structure behind the events described by a sentence, the kind of spatial representations they capture are representations in logical space, and define possible relations between objects and how they are represented in this logical space. I shall offer the precise logical details of this enriched logical space in the next section, in the proposal I shall call Visual Representation Theory.

3.4. A Logic of Vision, Part II: A Model of Visual Logical Space

In the previous section, we have been able to define a richer notion of visual object (i.e. things and their spatio-temporal properties), as well as sketching the nature of the relations holding between these objects. I shall integrate these results in our logic of vision as follows.

First, I shall assume that the set V of visual objects is now made of 'structured' entities, the combination of events e, objects o, and properties pr. The complex visual object that is made of these elements is the triple $v=\langle e,o,pr\rangle$, a basic entity which I shall call *Visual Representation Structure (VRS)*. Importantly, the set of events E (with $e\leq E$) is disjointed from that of objects O (with $o\leq O$) and the union of the two sets forms the whole set of (basic) visual objects, i.e. $E\cap O=0$ and $E\cup O=V$. Properties pr form a set of properties by which these visual objects can be individuated, i.e. we have $pr\leq PR$. In words, VRSs are made of basic objects (e.g. 'x'), the properties by which we individuate them (e.g. "swan"), and the events in which these properties are instantiated, i.e. their position in logical space with respect to other events. The following will hold: $v\leq V$, i.e. each VRS is part of the set of VRSs. I shall represent a VRS as $e{:}pr(o)$, which reads: An event e instantiates a property pr of an object o. This format follows the format of DRT, which I shall introduce in full in section 3.5.

We thus have seen that VRSs can be combined together via *merge*. The sum of two VRSs can be seen as a complex, novel event in which different properties of the same object can be combined together into a more complex, novel property. If $e{:}grab(x)$ and $h{:}order(x)$ are respectively an event of (pillow) grabbing and (pillow) ordering, then the complex event of (pillow) clean up can be formally defined as: ($e{:}grab(x)+h{:}order(x))=i{:}clean\text{-}up(x)$).

The structural properties of *merge* (associativity, commutativity, idempotence) are defined over VRSs as well, although the apparent 'temporal' nature of VRSs as representing 'objects in motion' requires some discussion. I shall focus on events, to make the discussion simple. An event of (pillow) clean-up can be organized in different ways (associativity); while we usually first grab pillows and then order them, when we clean up, an event of (pillow) clean-up consists of both events, regardless of their linear order (commutavity); several events of pillow-grabbing are still a (complex) event of pillow-grabbing (idempotence). Although VRSs are more complex objects, their 'combination' can be nevertheless defined via one basic operation, that of *merge*, which represents how complex VRSs are created from the union of basic VRSs.

The *part-of* relation is also defined over VRSs and events, and allows to define how events are structured. Reflexivity and transitivity allow to establish order/overlap among complex sequences of VRSs, straightforwardly enough. Antisymmetry allows to establish whether two VRSs (or parts thereof) are really the same, and thus to establish the identity between a complex VRS and the sum of its constituting VRSs. It also allows us to reconstruct their consequential/ temporal relation as well: if $e{:}grab(x)\leq i{:}clean\text{-}up(x)$, then $e{:}grab(x)\cup i{:}clean\text{-}up(x)=i{:}clean\text{-}up(x)$ and $e{:}grab(x)\cap i{:}clean\text{-}up(x)=e{:}grab(x)$. Since an event of pillow-grabbing is a proper part of (pillow) cleaning up, then it must precede the realization of a cleaning up event. The structural relations between events thus represent their causal/temporal relations: 'New' events come into being as the result of 'old' events being combined together, in an incremental fashion. If we don't grab and order pillows, we won't have an event of pillow-cleaning up: The existence of this event is a consequence of the combination of previous events.

The tight relation between the syntax and semantics of our logic of vision thus allows us to capture one aspect of 'dynamic' space by simply looking at how events are computed, without introducing further principles of analysis. Our new logic of vision can be thus represented as $S=<V,+,\leq>$, with V being a short-hand for $V=<E,O,PR>$.

This new logic of vision is a fully dynamic logic of vision when combined with an index set I, i.e. when we have $S_d=<I,V>$, with I being the index structure $I=<t,+>$. It allows us to explicitly represent how we integrate VRS together, one example being the following:

(21)	*t.*	$e{:}grab(x)$	(VRS instantiation)
	t+1.	$h{:}order(x)$	(VRS instantiation)
	t+2.	$e{:}grab(x)+h{:}order(x)$	(*merge* introduction)
	t+3.	$e{:}grab(x)+h{:}order(x) =i{:}clean\text{-}up(x)$	(sum of events)
	t+4.	$e{:}grab(x)\leq i{:}clean\text{-}up(x)$	(*part-of* rel. intr.)

In words, the merging of two VRSs yields a more complex VRS as the result, and allows to establish structural relations between VRSs. As we can see, the use of dynamic binding also allows us to bring out one aspect of the temporal nature of events: If we grab a pillow at a time t and then put it in order at a time $t+1$, then the resulting pillow-cleaning up event will be realized as a later time $t+3$, in a progressive way.

At this point, we have a quite rich and thorough *logic of vision* which allows us to model spatial representations/relations in a rather elegant and simple way, and which turns out to be somewhat similar to similar other logical theories proposed in, for example, the AI literature (e.g. *Event Calculus*; see Hamm & van Lambalgen 2005 and references therein and van der Does & van Lambalgen 2000 and e.g. Barwise & Seligman 1997 for non-linguistic applications of situation semantics). One example of the elegance behind our logic is the notion of ‘location’. VRSs explicitly represent the spatio-temporal ‘location’ of some event and its participants by representing the properties that individuate these entities. Geometric or mechanical properties are not any different from ‘grabbing’ properties, with respect to how this process occurs over time: we can thus represent e.g., the notion of inclusion as the VRS $e{:}in(x)$, that of support as $e{:}on(x)$ and so on.

We can then represent the notion of ‘motion’, or more appropriately the notion of change, as an ordering (*part-of*) relation between VRSs and the events they represent. So, if Mario goes in direction of the room and then stops once he’s in the room, he will be in the room as a consequence of this event of motion. This can be represented as $e{:}go(r)<i{:}in(r)$, i.e. an event of going into the room as expressing the relation holding between one event and its consequence. In a scenario in which Mario is sitting near the patio instead, other events may be going on at the same time, but at least these properties allow us to individuate Mario. We can represent this as $n{:}near(p)\leq z{:}gen(p)$, an event of (sitting) near the patio as part of a more generic event.

These relations between VRSs and the events they represent may find their way into language chiefly as adpositions via the function f, the isomorphism between vision and language. I shall re-define this function as follows. If f takes a pair of visual object and property as an input, it will return a noun as an output — we have $f(<o,pr>)=\boldsymbol{n}'$. If *f* takes a pair of event and property as an input, it will return a verb as an output — we have $f(<e,pr>)=\boldsymbol{v}'$**.** If it takes a full *VRS* as an input, it will return an adposition as a result — we have $f(<e,o,pr>)=\boldsymbol{p}'$**.**

The intuition is that 'partial' VRSs find their way in language as (common) nouns, labels for objects, and as verbs, labels for 'actions'; both individuate some entities, but do not express relations between these entities. Adpositions instead express the structural relations between VRSs, ultimately complex VRSs. The intuition is simple: Nouns (and verbs) find objects in logical space; adpositions denote the relations between these objects, which in turn represent a very abstract notion of space. L&J's syntactic proposals are still maintained, to an extent.[12]

The function *f,* as an isomorphism, preserves structure on VRSs as well: An adposition like *between*, for instance, is usually analyzed as the 'sum' of two simpler adpositions, such as *to the left or to the right of* some ground (e.g. Zwarts & Winter 2000). This can be represented as $f(r\text{-}of+l\text{-}of)=f(r\text{-}of)+f(l\text{-}of)$, i.e. the adposition representing the "between" relation is lexically equivalent with the adpositions representing the relations "to the left of" and "to the right of". Generalizing a bit, from basic spatial representations we can build more complex spatial relations; the complex structure defined by this process, the model of logical space defined by our logic of vision, may be represented in language *up to isomorphism*, via process of progressive refinement and specificity of relations (cf. also Levinson & Meira 2003). Hence, the mapping function *f* may assign different labels to its outputs, depending on the level of fine-grainedness and with some consistent cross-linguistic variation, but in a quite fine-grained and structurally regular way (cf. again Talmy 2000; see also section 4.2). Again, via this function we represent the possibility that we can match for each VRS a corresponding linguistic unit, and that the structural or 'spatial' relations between VRSs can find their way into language, chiefly as adpositions, at least in a language such as English.

Before moving to language, however, I shall make one observation regarding the nature of this process. According to the HCM proposal, when we mentally represent visual objects, these objects can be seen as output to some previous visual, perceptual input, which is then transduced as a visual object. This process occurs over discrete intervals of time, which in turn may be seen as minimal cycles of the *θ-rhythm*, and which may actually occur *independently* of the presence of external stimuli. In the absence of external stimuli, our brain still partitions the perceptual stream into minimal, discrete units. Very informally, our vision faculty will organize the perceptual stream into minimal units even if we are not observing finite objects such as tables, or if we look at the same portion of sky for quite a long interval of time.

When external stimuli are tracked, then it is possible to check whether they stand for some 'new' or 'old' information, i.e. whether their internal representation matches previous visual computations. Hence, the underlying properties

12 In this paper, I shall not propose an explanation on *why* the function *f* seems to operate such distinctions in the labeling process, and leave such a complex topic for future research.

of these computations do not crucially hinge on external stimuli, but on the possibility (perhaps, necessity) to integrate these different forms of information together in an effortless way, and in a coherent, 'synchronized' model (e.g., O'Keefe 2003, Buzśaki 2006). Our logic of vision thus represents an internal model of logical space, and represents the properties and relations defined over this model. By this point, our discussion of the logic of vision should be thorough enough: I shall concentrate on spatial language and its logic.

3.5. A Logic of Language: Discourse Representation Theory and Space

The study of meaning in natural language as a psychological phenomenon has long been adversed in model-theoretic approaches, traditionally rooted in an 'anti-psychologist' philosophy (e.g. Davidson 1967, Montague 1973, Cresswell 1985). Some modern research, however, broke with this tradition and attempted to study whether the models defined in this approach can be seen as mental structures and processes of some sort, represented via *dynamic logic* (e.g. Kamp 1981, Heim 1982, Chierchia & Turner 1988, Kamp & Reyle 1993, Chierchia 1995).

Among these different approaches, *Discourse Representation Theory* (DRT) represents the most important theory with such a 'cognitive' stance, and offers a set of tools which will allow us to (easily) treat all the linguistic phenomena we shall address via a single set of formal tools. For instance, it includes detailed treatments of the semantics of nouns and temporal expressions, which can be extended to treat our adpositions data (e.g., theories of noun reference and plurality such as Link 1983, 1998 or treatments of events such as Parsons 1990 and Landman 2000, 2004). It also allows us to take a perspective to sentence interpretation as a dynamic process, since it aims to model how sentences are interpreted and 'used' to form models in a compositional and incremental and on-line fashion, as in models of parsing such as Crain & Steedman (1985).

The version I shall use here is also fully compositional and thus allows us to analyze the contribution of each word to a sentence (iKamp *et al.* 2005, based on Muskens 1996 and van Eijck & Kamp 1997), and may be ideally implemented with certain minimalist theories of syntax with a 'processing stance' (parser-is-grammar of Phillips 1996). However, I shall focus on the contribution of nouns and adpositions for the most part, being somewhat sloppy on other parts of speech (such as verbs). Although the structural equivalences with my logic of vision should be immediately obvious, I will defer a thorough discussion to section 4.1, and focus here on the linguistic bits.

The most basic bits of information in DRT are Discourse Representation Structures (DRSs). A DRS can be thought, at a minimum, as a linguistic information state containing a set of discourse referents (or *U* for universe), an 'object' in discourse, and the conditions (or *CON*) which allow us to individuate such objects in discourse. While basic ('extensional') DRSs are at a minimum a duple of discourse referents (or individuals, for the sake of clarity) and their associated conditions, they 'become' information states when a third set of objects is taken in consideration, possible worlds (the set *W*). Hence, a DRS or information state is the triple $<W,U,CON>$ or $<\{w\},\{x\},\{\mathbf{con'}(x)\}>$, in which a discourse referent is paired with a 'world' referent and a condition, and which can be seen as a mental representation or (mini-)model that a speaker entertains, when he parses chunks of sentences, incrementally.

The nature of this world 'coordinate' deserves a few words of discussion. In classical logic, possible worlds are seen as quite real Leibnizian entities, such as the world we live in (e.g. Lewis 1986). Many versions of DRT, however, propose a different approach, partially based on Stalnaker's (1973, 1999) work,[13] in which possible worlds are mental objects, and represent nothing else than possible scenarios in which referents are involved, those for instance expressed by a sentence or a more complex text. Consequently, possible worlds can vary in 'size' and structure, and may be intuitively related one another according to the same principles definable over individuals, DRSs or other model-theoretical objects, as assumed in situations semantics (e.g., Barwise & Etchemendy 1990, Kratzer 1989, 2007) or modern modal logic (Hughes & Cresswell 1996, Blackburn *et al.* 2006).

Let us now turn to formal matters. As a standard convention, I write conditions in boldfaced characters and by adding a prime, i.e. '**con**′'. Hence, conditions in DRT are roughly equivalent to non-logical constants of first-order logic, and thus they represent 'concepts' or 'thoughts' as they are expressed in natural language, together with the distinction between *intension* and *extension* (cf. Margolis & Laurence 1999, Gärdenfors 2000, Winter 2008). The obvious consequence of this assumption is that our concepts/conditions will thus be invariably complex and definable in terms of their internal structure, unlike assumed in atomistic theories of concepts such as Fodor (1998, 2003). While an interesting topic per se, its discussion would lead us too far afield from our main topic of discussion, so I shall leave it aside for the time being.[14]

For our purposes, worlds and eventualities (i.e. events, properties changing over time, and states, properties holding over time) are basically the same (model-theoretic) objects, as in some variants of situation semantics. Very informally, if individuals represent objects, then eventualities represent the relations in which individuals are involved.[15] I shall use the term 'events' and avoid making any distinction between events and states, for the sake of clarity.

Once I have defined the basic structures of DRSs, I shall focus on the combinatorial and interpretative apparatus, i.e. how DRSs can be used to represent linguistic expressions. Here I shall use a variant of the 'linear' notation, rather than the more popular 'box' format, to enhance readability (as in Geurts 1999). I shall roughly match one syntactic phrase with one *DRS*, although more precise analyses are possible (see Kamp *et al.* 2005 for discussion). Look at the example:

---

13 This is true insofar as we look at the 'raw mechanics' of the underlying logic. Stalnaker's position is not a mentalist/internalist one: For him, 'possible worlds' are those of classical logic. DRT offers a much stronger mentalist perspective. Very informally, 'worlds' in DRT are roughly equivalent to possible thoughts or beliefs, information states ascribed to (thinking) agents. See Maier (2006: chap. 1) for discussion.

14 I would like to thank an anonymous reviewer for bringing this topic to my attention.

15 Note that, informally speaking, events and states are included in intervals of time, within the DRT architecture, with intervals of time forming up the main 'temporal structure' of a discourse. I shall diverge from DRT and use intervals of time in a different way, as I shall show in the remainder of the section.

(22) A man walks quickly. He whistles.

When a sentence like (22) is parsed, the parser builds up a bottom-up, left-to-right syntactic representation and, for each constituent and phrase, it builds up the corresponding DRS. For instance, *a man* is parsed as noun phrase/determiner phrase, and interpreted as the DRS [x:**man′**(x)], a DRS representing a referent x and a condition individuating him.

The next step consists in combining the predicate *walks* with the noun phrase *a man*. This is obtained via the syntactic operation *merge*, which shall represent as '+'.[16] *Merge* in DRT is a binary (associative, commutative, idempotent) operation that takes two DRSs and gives a 'bigger' (or new) DRS as the output, by unifying the universes and conditions of each DRS. In more formal terms, we have:

(23) [{x}:**con′**(x)]+[{y}:**con′**(y)]=[{x,y}:**con′**(x),**con′**(y)] (*merge* introduction)

In words, the merging of two DRSs forms a bigger DRS in which the universes and the conditions are merged pair-wise. Merged conditions are interpreted as being conjoined. If we were to translate conditions from our DRT language to first order logic, merged conditions would be interpreted as being conjoined, whereas each referent in the universe of discourse can be translated as an existentially quantified variable. We would have "∃x∃y[**con′**(x)&**con′**(y)]" for the two conditions in (23) (cf. Kamp *et al.* 2005: 143–145). I shall use brackets to mark the universe, and thus enhance readability (e.g. *{x,y}*), as in van Eijck & Kamp 1997 and Kamp *et al.* (2005).

The verb *walks* can now be simply represented as [e:**walk′**(x)], i.e. a DRS which introduces no new (object) referents but a novel spatio-temporal referent, the event of walking. The merging of the two resulting DRS can be represented, in a piece-meal fashion, as:

(24) *t.* [{x}:**man′**(x)]+[{e}:e:**walk′**(x)]=[{e,x}:**man′**(x), e:**walk′**(x)] (*merge* intr.)
t+1. [{e,x}:**man′**(x),e:**walk′**(x)]+[{e}:**quickly′**(e)]= (*merge* intr.)
t+2. [{e,x}:**man′**(x),e:**walk′**(x), e:**quickly′**(e)]

In words, we obtain the DRS representing the first sentence in (22) (*A man walks quickly*), by merging the DRSs representing its main constituting phrases. The DRS for *a man* acts as the context DRS, which is then updated via *merge* by the DRS for *walks*, acting as the context change potential DRS. The dynamic aspect of meaning is thus represented by the ability for new phrases/words to add more information regarding referents and events represented by each sentence, and thus define a 'broader' model representing facts. This is also represented via the explicit use of an index set in the derivations, which allows to explicit represent

[16] Kamp *et al.* (2005) use a different symbol, but this difference is immaterial, for our purposes. Note also that the properties of merge (associativity, commutativity, idempotence) stem from its definition as a (complex) form of set union, with idempotence allowing to 'reduce' universes whenever they are identical (see e.g. (22), i.e. *{e,x}+{e,x}={e,x}*).

how DRSs are combined together (as in Muskens 1996 and van Eijck & Kamp 1997, for example).[17]

The merging of DRSs has also one important consequence: it defines a semantic level of relations between DRSs and their universes/conditions, the *accessibility/part-of* relation. The *accessibility/part-of* relation is a transitive, anti-symmetric, reflective relation which allows to define one DRS *d* as part of another DRS *d'*, i.e. $d \leq d'$. While transitivity and reflexivity intuitively define how DRSs are connected over the flow of discourse, antisymmetry allows to make establish what relation holds between two referents/events/DRSs. One example is pronoun resolution: Intuitively, a pronoun such as *he* in (20) denotes one whistling individual as being a specific individual out of those who are walking quickly in the park. If at least part of the content expressed by two DRSs can be the same, then the two DRSs individuate the same object, a condition which expresses an anaphoric relation and is usually represented as $x=y$.[18] When the *accessibility* relation is restrained to discourse referents or events, it is usually called *part-of* relation (e.g. Kamp *et al.* 2005: 135). Consequently, I shall just use the *part-of* label for a semantic relation holding between DRSs, in order to make the exposition of the arguments clearer.

This is shown in the remainder of the derivation for (20):

(25) *t+3.* [{e,x}:**man'**(x),e:**walk'**(x),e:**quickly'**(x)]+[{e,y}:y=?,e:**whistle'**(y)]=
t+4. [{e,x,y}:**man'**(x),e:**walk'**(x),e:**quickly'**(e),y=x,e:**whistle'**(y)]

In words, the merging of the first and second sentence will also establish an identity relation between first walking man and second whistling man: There is really one man we are talking about, in (22). The resolution of the open anaphoric relation (i.e. *x=?*) amounts to identifying two referents by stating that the properties by which these referents are individuated converge to the same result.

After this brief introduction to the relevant aspect of DRT, I shall focus on a compact treatment of adpositions, which diverges from the standard DRT treatment of this category (cf. Kamp *et al.* 2005: chap. 2–3) and introduce a more thorough analysis of these terms, based on the vast literature on the topic. My basic assumption will match the non-linguistic considerations I offered in the previous section: Adpositions denote relations between DRSs, by expressing how the events denoted by these relations are ordered (e.g., Kamp 1979a, 1979b, Jackendoff 1983, Parsons 1990, Wunderlich 1991, Nam 1995, Fong 1997, Kracht 2002, Landman 2004, Zwarts 2005, Svenonius 2006, Ramchand 2008, and Kratzer, to appear).

---

17 In the dynamic semantics literature, the notion of 'dynamic binding' has a more restricted (semantic) application, and it is restricted to inter-sentential *merge*, i.e. the binding of information units over the sentence boundary (e.g., Chierchia 1995, Stockhof *et al.* 1996).

18 Pronoun resolution is sensible to features, like gender and number or temporal/aspectual values. I just ignore these aspects here, for the sake of clarity. In DRT, pronoun resolution also involves *presupposition* resolution, what could be (very) informally defined as the integration of implicit information in a DRS, together with the resolution of the anaphoric relations associated with this implicit information. See, among others, van der Sandt (1988, 1992), Geurts (1999), and Kamp *et al.* (2005: chap. 1–2) for discussion and references on this very complex and rich topic.

I shall thus assume that adpositions denote anaphoric relations between events/DRSs. Differently from pronoun anaphora, though, they may express 'asymmetric' relations, i.e. relations in which events are not necessarily identical. In this perspective, adpositions are akin to the 'duplex conditions' of DRT, which are used to represent quantifiers such as *every*, but also conditionals (e.g., donkey sentences), temporal adverbs and other temporal/logical relations.

The main reason for this assumption can be motivated by the following entailment patterns in the examples (adapted from Parsons 1990):

(26) A delegate walked into the park. → A delegate was in the park.

(27) A delegate is near the park. → A delegate is near the park.

In (26), the sentence *A delegate…* entails that the relevant delegate was in the park as a consequence of this event of motion. In (27), the sentence *A delegate…* entails itself, in the sense that it the delegate's position is not an explicit consequence of some previous event of motion, but also holds for possibly more specific states (e.g., the delegate being currently near the park). The symbol '→' represents the entailment relation between the two pairs of sentences.

The intuition behind these patterns is simple: adpositions, as they mirror relations between VRSs in language, also denote equivalent relations between DRSs and the events included in these DRSs. They do so by explicitly stating how events are ordered one another, thus explicitly representing the causal/temporal structure of (parts of) a sentence, possibly restricting this relation to certain events (e.g., those being "in" the park). I shall thus translate *into* as the complex DRS [{e,s,x,y}:*e<s*,s:**in'**(x,y)] and *near* as the complex DRS [{e,s,x,y}:*e≤s*,e: **near'**(x,y)]. The *DRSs* represent in a compact manner the Parsonian entailments, as *part-of* relations between the events denoted by the merged sentences. Informally, if a delegate walked into the park, he was in the park as a consequence. If a delegate is near the park, he may have arrived there because of some other events, or may stay there for some unspecified interval of time.

The interpretation of (24), at the relevant step and abstracting away from tense, is the following:

(28) *t.* [{e,x}:**delegate'**(x),e:**walk'**(x)]+[{e,s,y}: *e<s*,s:**in'**(x,y),**park'**(y)]=
t+1. [{e,s,x,y}:**delegate'**(x),e:**walk'**(x),*e<s*,s:**in'**(x,y),**park'**(y)]

In words, (28) says that a delegate walked and, as a consequence of this event of walking, he ended up in the park. The interpretation of (23) would be similar, except that the contribution of "near" would yield the following (slightly informal) *DRS*: [{e,s,x,y}:**delegate'**(x),s:**be'**(x),*s≤s'*,s:**near'**(x,y),**park'**(y)].

This treatment of English adpositions is by no means exhaustive and would probably need revisions, especially once we take in account a broader cross-linguistic perspective and the well-known interplay of adpositions and verbs of motion (again, see e.g. Talmy 1978, 2000, Svenonius 2006, Higginbotham 2009, and Zwarts 2010 for discussion). However, it allows us to represent in a rather simple what kind of contribution adpositions (and nouns) offer to a sentence, as well as introducing a rather compact theory of linguistic representation, in the guise of DRT. I shall thus collect all the crucial aspects of DRT and present them as parts of DRT's underlying logic.

DRT can be treated as a logic of language, which can be represented as the model $L=<D,+,\leq>$. The set D of DRSs is in turn a set of triples, defined as $d=<w,u,\mathbf{con}>$, and with $d\leq D$ holding for each d. The *model* (of discourse) defined by DRT is a *lattice* which has a structure entirely equivalent to that defined for vision.[19] The 'dynamic' incarnation of this model is $L_d=<I,D>$, the duple formed by DRSs and intervals of time at which they are combined together, with I again being defined as $I=<t,+>$.

The mapping from this model of language to other models, most specifically our logic of vision, can be easily defined via the function g, which is usually known as the *anchor function* in DRT (Kamp *et al.* 2005: chap. 4, Maier 2006: chap. 3 for discussion). This function is defined as an isomorphism mapping each linguistic information unit onto a non-linguistic unit, in this case a visual unit, i.e. $g(\mathbf{d'})=v$: In our case, it matches DRSs (linguistic information units) with VRSs (non-linguistic, visual information units).

Since it is an isomorphism, it maps at least one DRS onto one VRS, and at most one DRS onto one VRS. It preserves structure, so a mini-discourse like (22) can be seen as the description of a complex scenario, made of two connected, simpler scenarios. Formally, we have $g(\mathbf{d'}+\mathbf{k'})=g(\mathbf{d'})+g(\mathbf{k'})$, which in words says: The scenario corresponding to the mini-discourse in (22) corresponds to the scenario matched by the first sentence (a man is walking in the park) followed by the scenario matched by the second sentence (this man is whistling). Much like the function f, the function g can, but needs not to, find a VRS for each mapped term. In this regard, the function g can also be thought as representing a top-down process, since it represents how we can consciously match a sentence (and its content) with an extra-linguistic scenario it refers to.

Now that both sides of the isomorphism are defined, we have a good understanding of how information flows from vision to language and from language to vision, and thus we are ready to tackle the problem of the vision–language interface in an explicit way. However, before doing so, I shall offer an answer to the second research question, which is now within our reach. The answer is the following:

A–2: *Our models of spatial vision and language must include any possible property and relation that can 'connect' two entities; these models can (must) be treated via a model-theoretic approach.*

Spatial vision and language, then, can be seen as two systems representing different aspects of the same underlying phenomena: How we build up and maintain complex 'maps' of the objects we keep track of, over discourse. At this point, we can explore the common space generated by these two structures, and

[19] In DRT or similar approaches (e.g., Krifka 1998), events and referents are part of (structurally) different structures; here I follow Link (1983, 1998) and assume one common type of structure for all types of object.

thus focus on the vision–language interface.

4. A Theory of the Vision–Language Interface, and Beyond

In this section I shall offer a logical theory of the vision–language interface based on the results of the previous section (section 4.1); I shall offer empirical evidence in support of this approach (section 4.2); and sketch some broad consequences of my approach with respect to theories of the language faculty (section 4.3).

4.1. The Vision–Language Interface: A Formal Approach

A theory of the vision–language interface, given the discussion so far, must be a theory about the two-way information flow between two structures which represent (external) spatial information in a principled and highly organized way, the logical space defined by the logic of vision and language. As section 3 constituted a relatively long analysis of how these notions emerge from the basic bits of vision and language, I shall re-state my basic assumptions first and then dive into the vision–language interface problem.

I have assumed that both vision and language can be represented via a precise logic, which I called the logic of vision and the logic of language, respectively — or Visual Representation Theory (VRT) and Discourse Representation Theory (DRT), equivalently. These logical calculi share the same underlying structure: VRT is defined as triple $S=<V,+,\leq>$ and DRT as the triple $L=<D,+,\leq>$. These models are lattices, partially ordered sets, which minimally differ in having different types of elements, rather than in their structure.

The basic elements in these domains are respectively VRSs and DRSs: for each VRS v, the relation $v\leq V$ holds; for each DRS d, the relation $d\leq D$ holds. For each VRS v, the following identity holds: $v=<e,o,pr>$, i.e. each *VRS* is a triple of an event, an object and a property that identifies an object in an event. For each DRS d, the following identity holds: $d=<w,u,con>$, i.e. each DRS is a triple of a world/event, a referent and a condition that identifies a referent in a world/event. While VRSs are discrete units (possibly) representing perceptual stimuli from the visual apparatus, via transduction, DRSs may be seen as discrete units representing other types of information units (e.g., 'concepts' or 'thoughts'). They may be connected to VRSs via a slightly different type of transduction, but do not have a direct 'external' grounding: They represent purely 'internal' information.

While the two structures have different elements, their operations and relations are basically the same. A syntactic operation, *merge* (ultimately, set union), allows to define each element as the sum of other elements, possibly only itself. We represent it as '+'. Its definition is simple: It is a binary operation taking two inputs of the same type (e.g., DRSs: $a+b$), yielding an output of the same type as the inputs (a DRS: $a+b=c$). It is *associative*, *commutative*, and *idempotent*: It allows to combine the same elements in different ways (associativity: $(a+b)+c=a+(b+c)$), regardless of their order of occurrence (commutativity: $a+b=b+a$), and can be 'repeated' on the same input (idempotence: $a+a=a$).

A semantic relation, the *accessibility/part-of* relation (represented as '≤'), integrates this syntactic operation and establishes how the results of the *merge* operation are 'connected'. It is binary, as it establishes a relation between two objects of the same type (e.g., VRSs: $a \leq b$, and it is reflexive, asymmetric and transitive: It allows us to establish that objects are part of themselves (i.e. $a \leq a$), that objects can be identified (i.e. if $a \leq b$ and $b \leq a$, then $a=b$), and that multiple relations can be compressed as a single relation (i.e. if $a \leq b$ and $b \leq c$, then $a \leq c$).

The *merge* operation and the *part-of* relation are connected via the following properties, which I shall again represent via set-theoretic notation. If $a \leq b$, then $a \cup b=b$ and $a \cap b=a$. In words, if one object is part of another, then their merging will correspond with the 'bigger' object (union), and their product will correspond to the 'smaller' object (intersection). Semantic relations can be seen as the result of previous instances of syntactic operations, in a sense recording the successful merge of two objects into a more complex, novel object. The structures defined by these operations are complex Lattices, i.e. partially ordered sets with a syntax and a corresponding semantics, and thus models of the phenomena they represent.

Although other operations can be defined (e.g., set intersection standing for attention), this 'minimal' logic allows us to aptly model how information units are processed and integrated together into more complex units, in a bottom-up way. They also allow us to define how one logic can be tightly connected to another via two functions, f and g, which respectively define an isomorphic mapping from VRSs to DRSs and from DRSs to VRSs. These functions are isomorphic because they map at least one input and at most one input to the same output, i.e. they are respectively injective and surjective, thus they are bijective.

The function f is defined as: $f{:}v \rightarrow d$, i.e. a function that maps each visual structure $v \leq V$ onto a discourse structure $d \leq D$, whereas the function g is defined as: $g{:}d \rightarrow v$, i.e. a function that maps each discourse structure $d \leq D$ onto a visual structure $v \leq V$. Note, now, that these two functions are one the inverse of the other: Their *composition* (represented via the symbol '$\circ$') will yield the identity function, e.g., we have $f \circ g=i$, with 'i' being the *identity function*. This latter property tells us that, for example, each noun may act as the linguistic label for a visual object, and thus that each visual object may a have noun as a linguistic label.

These isomorphisms allow us to explicitly represent how we 'translate' one type of objects into another, while for logical operators (i.e. *merge* and the *part-of* relation), they offer evidence that these operations are the same across models/ logical systems. The reason is simple: while objects define non-logical constants, *merge and the part-of* relation define logical constants, elements of a logic that receive the same interpretation on any model, whether it represents vision or language. In words, *merge* is interpreted as the union of two objects, whether these sets stand for visual structures or discourse structures, and so is the *part-of* relation interpreted as a relation between objects.

This is explicitly represented via a structure-preserving condition on our isomorphisms: $f(a+b)=\boldsymbol{a'}+\boldsymbol{b'}$, given that $f(a)=\boldsymbol{a'}$ and $f(b)=\boldsymbol{b'}$. In words, the noun for the object corresponding to the merge of the object "legs" and "surface" (*table*) corresponds to the super-ordinate noun that stands for the objects "legs" and "surface". The *merge* symbol is the same on both sides of the identity, while the merged objects are different. The same holds for the *part-of* relation, since, if we have $a \leq b$, we have $f(a) \leq f(b)$. In words, if a leg is part of a table, then the noun/concept *leg* is part of the noun/concept *table*. The same considerations hold, *mutatis mutandis*, for the function g. In words, vision and language may differ as models representing different 'things', but they are equivalent as models sharing the same structure.

The definition of these two isomorphisms has one important consequence: It allows us to outline a simple and yet very precise theory of the vision–language interface. The main assumption I shall make is that the vision–language interface is defined as a *Galois connection* between these two structures. A Galois connection is defined as follows: given two lattices $<A,\leq>$ and $<B,\leq>$, $f(a) \leq b$ *if and only if* a≤g(b). In our case, and with some notational fantasy, given the lattices $<D,\leq>$ and $<V,\leq>$, we have $g(\mathbf{d'}) \leq v$ if and only if $\mathbf{d'} \leq f(v)$. In words, if vision and language are connected via a Galois connection, then the VRS corresponding to a DRS is part of a larger VRS, and a DRS corresponding to a VRS is part of a larger DRS. In words, vision and language representations are connected if each linguistic term is matched by a visual entity, which is part of a 'larger' scenario, and if each linguistic term expressing a visual object is part of a sentence. Informally, a Galois connection is a method of defining an isomorphism between structures in which weaker relations can also be defined: it allows us to express not only that structures 'look the same', but also to compare the relation between many elements of one structure to an element of the other structure (e.g., Ganter & Wille 1998: chap. 1).

The strength of this proposal is that it allows us to define a degree of accuracy by which a certain sentence describes a state of affairs and *vice versa*. For instance, an adposition matches a spatial representation when the two following conditions hold: $f(v)=\mathbf{d'}$ and $g(\mathbf{d'})=v$. In words, if a book is supported by the top vertical surface of a computer, then the adposition *on top of* is quite ideal match for this scenario, since we intuitively have *f(on-top)=**on-top'***, but also *g(**on-top'**)=on-top*.

While identity cases are in a sense trivial, cases of partial matches allow us to grasp the crucial strength of the proposal. For instance, an adposition expressing only support of the book by the computer is intuitively less accurate (i.e. *on*) than *on top of*, which expresses the specific surface offering this support. This because it will represent only a part of the spatial representation in which book and computer are involved: if *g(**on'**)=on* and *on≤on-top*, then we will have *g(**on'**)≤on-top* to hold. In words, *on* represents only a part of a certain extra-linguistic scenario, and thus will be less accurate than *on top of*. Conversely, the relation ***on'**≤f(on-top)* also holds, i.e. *on* is less accurate than the adposition which would perfectly match the said scenario. Hence, the *part-of* relation, when it is defined on 'mixed' objects by means of a Galois connection, can be interpreted as relation expressing a degree of accuracy of a sentence, an adposition or any part of speech, with respect to the extra-linguistic context.

This proposal on the vision–language interface makes two main predictions. First, it predicts that the 'amount' of spatial (visual) information expressed by a sentence is flexible, and may be as accurate as the corresponding visual scenario, but also that the same scenario can be described by adpositions of different 'accuracy'. Second, it predicts that, since the 'binding' between the two layers of information may go in both directions, there is no 'causal' relation between these different computations, so one type of information is processed *independently* of the other. We are quite able to evaluate whether what we see refers to (or matches with) what we say and *vice versa*, but both mental processes need not a constant, unconscious feedback between the two levels of comprehension to occur. In words, we can say a lot about 'where' things are (including, but not limited to, geometric relations), but need not to limit ourselves to what we see.

A formal treatment of this 'parallel' processing can be represented as follows:

(29)	I	V	$V \Leftrightarrow D$	D
	$t.$	$(a+b)$		$(\mathbf{a'}+\mathbf{b'})$
	$t+1.$	$(a+b)=g(\mathbf{a'+b'})$	$f(a+b)=\mathbf{a'+b'}$	
	$t+2.$	$(a+b)=g(\mathbf{a'+b'}) \Leftrightarrow f(a+b)=\mathbf{a'+b'}$		

In words, at some interval in a computation, the two types of information are first mapped onto the other domain, and then (dynamically) bound together if the two 'flows' of the process yield the same result, possibly compared in terms of accuracy in a common logical space, which is represented as '$V \Leftrightarrow D$'. Informally, we check if what we see matches with what we say and *vice versa*, hence obtaining a 'broader' picture of facts. Since what we see needs not to match with what we say, the binding relation between these two types of information is entirely optional and, as we have discussed so far, it ultimately represents a top-down translation process, which can be more or less accurate.

One important thing to note is that this formal treatment is modular also because the binding of two types of information is explicitly represented as a distinct result of a matching operation. If we would have assumed that the binding occurs by the simple co-synchronous occurrence of these operations, our architecture would actually have been *connectionist*, in nature. While the two processes are isomorphic and can be tightly connected, they are nevertheless two distinct processes, and a third process is their matching relation (i.e. binding); see Marcus (2001) for discussion. Now that we have gone through the formal details and their predictions, we can focus on their empirical support, which I shall analyze in the next section.

4.2. Testing the Theory against the Data

The theory I have proposed in the previous section is consistent with general assumptions about vision and language as parts of a cognitive and modular architecture (cf. e.g. Jackendoff 1997, 2002), and possibly offers a more fine-grained and formally precision analysis and representation of these modules and their processes. In this section I shall explain more in detail why this theory is consistent with previous proposals and offer an 'improved' model of their insights, and why it is consistent with general assumptions about cognitive architecture, i.e. why the two main predictions I offered in the previous section hold. I shall focus on four topics, offering evidence that confirms these predictions.

A first topic pertains to the 'amount' of space found in language. Let me repeat (13) and (14) as (30) and (31) to illustrate the point:

(30) The book is on the tip of the left edge of the blue table.

(31) The book is on the table.

The crucial difference between (30) and (31) is that both sentences may be used to convey information about the same extra-linguistic scenario, but (31) is definitely more accurate than (30). Vision-wise, a scenario in which the book is supported by the tip (of the edge) of the table is also a scenario in which a book is supported by the table — hence, the relation *on≤on-top* holds. language-wise, the DRS representing (31) is part of the *DRS* representing (28), so the relation ***on'≤on-top'*** holds. Hence, the following identities *g(**on-tip'**)=on-top* and *g(**on'**)=on* hold, as well as ***on-tip'****=f(on-top)* and ***on'****=(on)*. We can then observe that the relation *g(**on'**)≤on-top* holds, i.e. that (31) is a partial representation of the same scenario that (30) is a total representation of, and thus a less accurate description of facts. Conversely, the relation ***on'****≤f(on-top)* holds, i.e. (31) expresses part of the information expressed by (30), and thus of the scenario that (30) represents.

A second topic pertains to the different degree of accuracy that two sentences can have in describing a certain scenario, when involving different adpositions. If the meaning of two adpositions overlaps or stands in an entailment relation, then speakers may favor one over another, when they need to associate it to visual information. The entailment cases are quite intuitive, and can be seen as a general case of the relation between (30) and (31). In a situation in which a book is supported by the upper part of a drawer, *on top of* may be judged as a 'perfect' adposition to describe this situation while *on*, that is entailed by *on top of*, may be considered as less appropriate, with respect to the scenario it purports to match with.

The cases in which adpositions overlap in meaning require some more discussion. Let me repeat (10) and (11) as (32) and (33) to illustrate the point:

(32) The painting is on the wall.

(33) The painting is in the wall.

In a scenario in which a panting is literally encased in the wall, (33) may be a more accurate sentence to describe this scenario than (32), because it may express in a more precise way the matching extra-linguistic scenario. Intuitively, if a painting is in the wall, it is certainly supported by it, and actually part of the wall's surface, rather than just adjacent to it (as for *on*). Formally, we can say that *in* is more accurate than *on* with respect to the aforementioned scenario if the following holds: if **in'≤*on'***, then ***on'*∩*in'*=*in'***, i.e. *in* is a part of *on* and its meaning; and thus, if *g(**on'**)≤g(**in'**)*, then *g(**on'**)∩g(**in'**)=g(**in'**)*, i.e. *in* describes a more specific scenario than *on*, and is hence considered more accurate.

The treatments I discussed in the first and second topic are consistent with results like those of Coventry & Garrod (2004), Regier *et al.* (2005), and much of the aforementioned literature on spatial sentence processing, which also cover the relations between e.g., *above* and *on*, *in* and *under*, and so on. It is also consistent with Levinson & Meira's (2003) cross-linguistic results, which are indeed based on how adpositions can be conceptually organised in terms of increasing accuracy and specificity of their use in (implicit) context.[20] This literature also offers indirect evidence of the validity of my proposal: Most experiments aim to test how participants consciously match visual stimuli with linguistic stimuli, evaluating how accurate sentences can be in describing a scenario. Hence, it indirectly supports the view that the functions f and g represent conscious processes.

A third topic pertains to a complex case, that of the relation between vision and language with respect to reference systems and their computation. Again, works like Carlson-Radvansky & Irwin (1994), Carlson (1999), or Regier *et al.* (2005) show that, when speakers interpret axial terms such as *to the left of*, their accuracy can be measured with respect to different reference frames, e.g., whether a chair is to the left of a table with respect to the observer (relative frame), the chair itself (intrinsic frame), or an environmental cue like the floor (absolute frame). What I have suggested for 'standard' adpositions can be extended to these 'axial' adpositions as well, with no need to make any further assumptions. Furthermore, although some proposals conjecture that the 'cognitive' procedures by which 'absolute' spatial relations are computed dramatically differ from other visual procedures (e.g. Levinson 2003), their mapping onto linguistic unit seems to be rather 'ordinary'. Whether we may compute a polar direction such as the one corresponding to *North* via an entirely different set of cognitive resources than the ones involved in e.g., computing the support relation corresponding to *on*, the two adpositions share the same underlying grammar, and seem not to reflect this 'cognitive difference', if it exists.

From these three topics we can observe that the first prediction of my novel interface approach, the flexibility of this interface, is substantially borne out. This allows to make a further general comment regarding the "how much space" problem, and how we may choose the degree of accuracy we want to express. The literature gives us the relevant answer regarding how this process comes about, in the guise of theories of sentence planning and production. For instance, in a theory of sentence-planning (*speaking*) like Levelt (1989), speakers are assumed to decide, at a pre-linguistic level, both which basic 'facts' and the relations between these facts they wish to convey (Levelt's level of *macro-planning*), and consequently which language-specific rules (syntactic and semantic alike) to use in order to convey these facts (Levelt's level of *micro-planning*).

For our discussion macro-planning represents the relevant aspect of production, since it indirectly defines "how much" we may express about extra-linguistic information. In slightly more formal terms, macro-planning may be treated in the following way. A speaker may look at a certain general visual context V and may decide to express part of this scenario via the selection of a certain VRS v. Given a selection function s, this process can be represented as, for

[20] A conjecture is that classical results of prototype theory (e.g. Rosch 1975) may actually find a formally precise account, if we, for example, pursue the intuition that a noun such as *robin* may be seen as the perfect linguistic label for the sum of all visual/cognitive information we ascribe to birds, rather than *penguin*. This intuition is actually pursued in Ganter & Wille (1998) and especially in van Eijck & Zwarts (2004) in thorough detail.

example, $s(V)=v$. For instance, a speaker may look around a room and may decide to say that a certain specific book is on the tip of the left edge of the blue table. The selected VRS v would actually stand for the complex VRSs representing book, blue table, edges, and tips, and the relations holding between these VRSs.

The sentence corresponding to this VRS, which we can represent as $f(v)=\mathbf{S'}$ and thus as $f(s(V))=\mathbf{S'}$, indirectly represents which pre-linguistic facts are chosen by the speaker as finding their way into language. The amount of space finding its way into language roughly corresponds to the speaker's intentions to be more or less accurate in describing a scenario and his eventual desire to express one outstanding aspect over another. Although he may do so via different microplans, i.e. via the choice of different words and sentence, this choice is inherently flexible, rather than dictated by constraints on what type of spatial information finds its way in language. This is captured by the function f taking the function s as its input. Informally, we may decide to say something about the scene we are paying attention to and, in doing so, we *selectively* (and consciously) pick out visual information about this scene, then 'convert' it into the corresponding sentence, thus effectively deciding how much 'space' gets into language.

A fourth topic pertains to the relation between vision and language in case of cognitive impairment in one of the two modules. The intuition is the following: If my theory can predict how the vision–language interface works, it should also make predictions about the problems that could arise when these modules are not properly interfaced — it should be *breakdown-compatible* (e.g. Grodzinsky 1990). The following examples suggest that this is indeed the case.

A well-known fact is that people affected by Williams syndrome may have relatively intact language skills, including a good understanding of spatial language, but are usually unable to assess even basic spatial relations from a visual perspective. These patients may be able to understand an adposition such as *in front of*, but may not be able to evaluate what is the front of an object (e.g. Landau & Hoffman 2005 and references therein).

An obvious account, in the current proposal, is that, since spatial vision is quite impaired, it will not be possible to have a visual input that will correspond to a linguistic output. That is to say, the function $f(v)$ will be undefined, since it will have no input, and so the function $g(\mathbf{d'})$ will be undefined as well. As a consequence, it may not be possible for individuals with Williams syndrome (to give one example) to relate what they see to what they say. As it stands, my proposal seems to be consistent not only with a general modular approach to cognition, but also with a general approach to cognition and its disorders.

Another well-known case of a cognitive disorder affecting one side of the 'space' interface is aphasia. In Broca's aphasia, omission of prepositions (among other functional words) is well attested, while spatial vision is usually (completely) spared. Adposition omission in aphasia may be gradual and patients tend to omit more general adpositions (e.g. *at*) rather than less general adpositions (e.g. *in front of*; see e.g. Trofimova 2009 for a recent review). Regardless of their degree of language impairment, aphasics usually lose their ability to produce but usually not their ability to comprehend adpositions, and, more generally, language; hence, they are able to understand whether adpositions correctly describe a scenario or not. While one aspect of spatial language can be dramatically impaired (e.g. production), all other aspects of both spatial vision and language, including their interface, are substantially spared, in line with my assumptions.

A similar account may be extended to another cognitive disorder, that of dyslexia.[21] Models like the *Dual Route Cascaded* model of reading aloud (*DRC*; e.g. Coltheart *et al.* 2001 but also Beaton 2004), the processing of ('reading') a single word is assumed to occur via three parallel processes, one in which we visually recognize a written word (*non-lexical route*), and one in which we (may) retrieve its lexical entry as well as its phonological and syntactc-semantic properties (*lexical/sub-lexical route*). Although one process can be faster than the other, full recognition of a word occurs when both processes converge to the same output, but fails if the 'visual' process is damaged (failure to read graphemes and words, or *shallow dyslexia*) or the 'linguistic' process is damaged (failure to understand the meaning of words, or *deep dyslexia*).

As per the other cognitive disorders, our theory of the vision–language interface is consistent with this analysis of dyslexia without any further assumptions. Although for dyslexia we would certainly need a more accurate and specific analysis of both sides of the problem, the intuition seems to be correct: We may not be able to see certain visual objects correctly, but we may still retrieve their corresponding linguistic labels, and *vice versa*. We can also observe that the second prediction is borne out, since these cognitive disorders show that spatial computations can occur both at the visual and linguistic level and can be bound together, but also that this binding process is not necessary. In fact, even if one side of this process may be completely impaired, the other side will be still able to work independently.

Summing up, the discussion of these four topics suggests that our vision–language interface theory can have theoretical value and can withstand empirical scrutiny, even once we look beyond the topic of space. As we have seen, visual and linguistic representations can be matched in a quite precise way, but the processes regulating this matching of information is inherently conscious, that is, based on a speaker's top-down thought processes. Speakers may wish to be more or less accurate in describing a scenario and may evaluate sentences with respect to their descriptive accuracy. They may be able to understand spatial language even if they can't navigate the environment and, for complex tasks such as reading (i.e. the codified matching of visual and linguistic stimuli), they require conscious and protracted effort to establish the proper mappings, provided that this mapping is not impaired by cognitive deficits.

These facts are somehow hard to explain in previous accounts of the vision–language interface but fall out as predictions of the theory I have sketched so far due to its flexibility. This theory also presents in detail the convergences between space, vision, and language, offering a view in which these two modules are remarkably similar; as such, it may appear that there is little or no difference between the two modules, both from a structural and a content-bound point of

21 Dyslexia can be informally defined as a cognitive disorder which influences our ability to successfully read, i.e. to either successfully decode the sequence of graphemes ('letters') making up a written word, or to properly interpret a word, and access syntactic information about it. See Beaton (2004) for a thorough introduction.

view. I shall focus on these differences in the next section.

4.3. *What Is Unique to Language, and Why*

The discussion I have offered so far has sketched the strong similarities between vision and language as modules of cognition. It has also offered an attempt to explain how these two modules exchange information — for instance, via the synchronization of their processes. (Spatial) vision and language seem to be remarkably similar modules, and it is not surprising that in some quarters they are considered as contiguous modules, if not the same module, in some respect (e.g. Talmy 2000, Coventry & Garrod 2004).

There are, however, a number of properties of language which seem rather hard to reduce to general, non-linguistic features, and which inherently involve the possibility in language to convey information about 'unbounded' quantities. In the common parlance of biolinguistic research, much of our discussion up to this point has focused on defining the properties that can be ascribed to the broad faculty of language (FLB), in the terminology of Hauser *et al.* (2002), since I have mostly been concerned with the relation between vision and language, and with those properties that are shared by both computational systems. In this section, I shall sketch a very preliminary proposal, stemming from the discussion offered so far, on what properties are unique to language and thus may be possible candidates to form the kernel of the faculty of language in the narrow sense (FLN). I shall do so by focusing, for the most part, on spatial language. I shall discuss these properties in a less formally rigorous way, focusing on speculative aspects of the discussion.

Look at the examples:

(34) Mario has gone to the store *three times*.

(35) Mario *may* go to the store.

(36) *All* the boys have gone towards the store.

(37) *Every* boy will go toward the fence.

(38) *A* boy may come to the party.

(39) *Some* boy may come to the party.

(40) Mario *always* goes to the store.

(41) Mario *seldom* goes to the store.

(42) *Where* are you going?

(43) I am going *there*, too.

(44) Mario *lends* a book to Luigi.

(45) Luigi *borrows* a book from Mario.

In (34), Mario's going to the store is described as occurring three times or instances, but little is said about when this happens: It may occur one time right now, one time yesterday, and one time when he was a young lad. Two of the events that the adverb denotes cannot be mapped onto visual inputs, because two of them cannot correspond to current facts, but rather to 'memory traces' we have recorded of them. Language allows us to merge together pieces of information which do not necessarily correspond to one modality, into a unified type of information.

In (35), Mario's possible event of going to the store is something that we conceive as occurring in, say, a few more minutes, or whenever he feels like it, perhaps tomorrow. In the case of the non-current events of (34), the modal auxiliary may simply denote a linguistic unit which hardly can find a visual unit as its counterpart. In (36) and (37), the amount of boys that have gone to the store may vary, and may involve pairs or triples (or bigger quantities), but each of these possible combinations of boys will go to the store, without any exceptions.

In (38) and (39), instead, we may not know the identity of who is going to come to the party, except that it is likely to be a single boy, someone who we may have not mentioned so far and may never come to know, let alone see. These cases may already show that the mapping from vision to language can be quite partial (i.e. not always defined), but the following cases should give even stronger evidence. Adverbs such as *always* and *seldom*, as in (40) and (41), suggest that we may even convey linguistic information about several (infinite) situations (sets of events) in which Mario goes to the store, or say that such situations are rare but do occur (i.e. *seldom*).

Examples like (42) and (43) show that we may actually rely on someone else's ability to access information in order to retrieve information of Mario's whereabouts: If someone answers our question, we will be able to know Mario's location without actually seeing this location, and if someone has already told us *where* Mario is going, we may say that we are going *there*, although we may not be able to see "where" "there" is. In (44) and (45), the same set of events is presented under two different, and in a sense complementary, perspectives: While the visual scenario is in a sense the same (a book is temporarily exchanged, between Mario and Luigi), the two sentences express these facts from Mario or Luigi's perspective, respectively.

There are two generalizations that we can make from these examples. One is that language may convey information which can be *multi-modal*, in the sense that linguistic units may bring and represent together information which comes from different cognitive sources, and may have no extra-linguistic instantiation whatsoever. This is not surprising if we look at language at a module that only processes internal information, stripped of any perceptual or modal-specific aspects (unlike vision), but it is also consistent with various theories of memory as a 'mental' model in which we record and organize memory.

One way to look at this aspect is the following, and it is based on theories of memory like Cowan's (1988, 1995, 2005). In this theory, long-term memory is seen as the model representing all the information we may have stored about the world, whether it is veridical or not (i.e. whether it is represented in episodic memory or not[22]). Short-term memory, on the other hand, can be seen as the current

---

22 Episodic memory is a component of memory which 'records' perceptual information regard-

part of long-term memory which is accessed and evaluated at a given time. In our logic, long-term memory can be seen as a static model <*D*> or <*V*>, while short-term memory can be seen as the dynamic counterparts of these models, <*I,D*> or <*I,V*>.

For instance, we may have observed Mario going to the store in three very different moments of our life, but if we use a sentence like (32), we represent these otherwise separate events of time in the same representation (ultimately, a DRS) in our short-term memory. Language allows us to define a 'common space' in which 'displaced' events form a may form a consistent representation insofar as they share the same formal properties (e.g., being three instances of a walking event), and thus are stripped of any constraints on perceptual information — but may also be bound with other 'portions' of short-term memory (e.g., visual computations; cf. the previous section). Informally, an adverb like *three times* says that there are three contiguous intervals in a derivation in which three events of going to the station become logically contiguous, i.e. we have $a+b+c$ at an interval $t+n$.

Another generalization is that language can express relations and quantities which are not necessarily finite (or bounded), and is not limited to offering one perspective. This latter, (quite) rough, intuition is based on our last pair of examples, but several other similar examples could be made; think of any active sentence and its passive counterpart, for instance. If we think in slightly more formal terms, we may construe (42) as representing a scenario in which Mario's actions as an 'agent' operates onto Luigi as a 'patient', and can be schematically represented as $a \rightarrow p$. We can then assume that (43) can be represented as the inverse type of relation, which can be represented as also $\neg(a \rightarrow p)$. In very informal words, we can represent that the sequence of events expressed by (43) flows in the opposite direction of (42), as the informal use of negation aims to represent, although we express the order of relevant entities in the same way as in (42).

This is possible because in language we can express the same underlying conceptual structures under different 'perspectives', but via virtually the same logical apparatus (cf. also Landman 1991: chap. 3 and 2004: chap. 7–8). Again, if we think of language as defining a conceptual 'space' not constrained by perceptual limits, then the same underlying information can be expressed in two apparently opposite ways, which, however, underlie the same logical principles (e.g., *monotonicity*). Although (42) and (43) describe the same extra-linguistic event, their interpretations represent two possible ways by which language can structure this information.

Another form of 'unboundedness', its linguistic realization as well, is ultimately represented by the interpretation of quantifiers and other 'expressions of quantity', as the examples show. Informally, in language we can express information about numbers of individuals which are far greater than the amount of individuals we can 'see' and which can be structured in rather complex and fine-grained ways, as adverbs like *seldom* and *always* suggest.

This can be illustrated via a detailed analysis of (34) and (35). Note here that I shall depart quite dramatically from DRT and treat a quantifier like *every* as represented by a logical operator rather than a duplex condition. In both senten-

ing the first time we observe a given event.

ces, it is possible to represent the contribution of *all* and *every* to the sentence in terms of the universal quantifier, which I shall here represent in its Boolean incarnation, '∧' (e.g., Montague 1973, Keenan & Faltz 1985). This symbol can be informally interpreted as a form of unbounded coordination: Informally, the sentence *Every boy has gone to the store* can be interpreted as the equivalent of "Mario has gone towards the store and Luigi has gone to the store a*nd*…", i.e. as if we were to state each possible boy in a large, perhaps infinite, domain of discourse, one by one.

Suppose then that we take the set of boys as a list (sequence) of boys in discourse. The DRS representing *all the boys* is equivalent to the merging of the DRS representing the sum of the last boy with the sequence of boys occurring before him in this infinite list. We define the interpretation of a universally quantified noun phrase (its DRS) via the sum of the interpretation of its parts, via induction (its constituting DRSs).

This can be represented as:

(46) *t.* $[\{x{-}2\}:\mathbf{boy'}(x{-}2)]+[\{x{-}1\}:\mathbf{boy'}(x{-}1)]=$
t+n. $[\{x{-}2,x{-}1\}:\mathbf{boy'}(x{-}2),\mathbf{boy'}(x{-}1)]=$
t+n+1. $[\wedge x:\mathbf{boy'}(x)]$

With the referent/individual *(x–2)* representing the list of boys preceding the last boy (i.e. the second-to last (complex) referent), *(x–1)* representing the last boy, and $\wedge x$ representing the 'new' referent obtained from the merging of the two 'old' referents. This is a *recursive*, *inductive* definition of the universal quantifier, in terms of an unbounded form of *merge*, and its interpretation. In words, we interpret the universal quantifier as the result of taking each referent in discourse via one common condition. This result is another DRS, the DRS representing the result of taking each referent which can be identified as a "boy" one by one, i.e. via the product of each condition merged in a DRS, here represented as '∧'.

These considerations can be also extended to other quantifiers with the proper *provisos*, but also to adpositions, thus suggesting that spatial language is also 'unbounded' in its interpretive range. For instance, the relational component of any adposition (e.g. *near*) can be recursively defined as the merging of two opportune relations. Abstracting away from the specific condition on proximity (i.e. **near'**) and with some notational fantasy, "near" can be represented as:

(47) $[\{s,s'\}:s\leq s']=[\{s,s'{-}2\}:s\leq(s'{-}2)]+[\{s,s{-}1\}:s\leq(s'{-}1)]$

Here the 'geometry' approach to adpositions is quite useful to illustrate the intuitive meaning of (47). If a figure is the ground when it occupies a certain region, then it will be near the ground if it occupies any region which is included in the bigger region. Conversely, once we sum all the (sub)-regions in which a figure is near a ground, then we will obtain the 'general' region which can be labeled as *near*.

This way of representing the universal quantifier, and in general of representing quantified noun phrases, as well as the interpretation of *near* and other adpositions, is informally based on one recursive function, the *Fibonacci series*, which allows to define one object (e.g., a natural number) as the sum of its direct predecessors. Intuitively, it may be extended to all of the other functional words I have discussed in examples (34)–(45), and to any expression that captures a form of quantification.

Several authors have argued that the Fibonacci series can represent how recursion is expressed in language (e.g. Soschen 2008 and references therein), but one may also assume that the successor function may be a recursive function that can also be used to represent the recursive nature of syntactic processes (see Landman 1991: chap. 1 for discussion). The crucial aspect is that, since language is different from vision by being fully recursive at a syntactic level, it will also be different in having terms which directly express the result (interpretation) of this unboundedness, and thus will be fully recursive at the semantic level.

An indirect way of capturing this difference is by enriching our logic representing language, so that we have the tuple $L=\langle D,+,\leq,\wedge\rangle$. This tuple represents the 'structure' of language as including not only a minimal syntax (*merge*, '+') and semantics (the *part-of* relation, '≤'), but also a set of operators, here represented by the universal quantifier, that denote *the result* of linguistic processes. We have an indirect reconstruction of the distinction between FLN and FLB.

This reconstruction is indirect only because recursion is a resulting property of the 'logic of language', but it nevertheless represents one (maybe *the*) element of distinction between vision and language. Informally, it tells us that language has certain recursive *closure principles* which allow to label not only objects from other models (e.g. nouns for objects), but also to express the processes by which we collect together these objects into abstract structures. Adpositions represent one case, and quantifiers represent a more language-specific case, but the same reasoning could be applied to any functional word in language. We are able to talk about, for example, *all the past boys and apples* because we are able to compute a referent that stands for the combination of two different sets of entities (i.e. boys and apples), possibly representing entities 'displaced' in time and space — even if these sets may include an infinite amount of 'smaller' referents (i.e. each single boy and apple).

It is also indirect because, as we have seen, visual and linguistic information are processed as distinct types of information, although they are potentially connected up to isomorphism. While there can be an intimate relation between what we see and what we say, language is not bound by other modules of cognition in its expressive power, although the entities that make up the universe of discourse denoted by language must be the result of previous processes of interpretation, as the closure principle entails.

One important aspect, however, is again that vision can be represented via a similar, although less 'powerful', logical structure: As observed in Pinker & Jackendoff (2005) and Jackendoff & Pinker (2005), a number of 'structural' or hierarchical properties are *domain-general*, and thus not unique to language, because they represent domain-general logical principles by which we process, retain, and organize different types of information. Vision represents here one important case, but the phonological component of language also offers a similar case, and other examples abound (e.g., the 'grammar of action' analyzed by Fujita

2009, the 'grammar of music' in Jackendoff & Lerdahl 2006, or the 'grammar of phonetics' of Reiss 2007).

The intuition behind these considerations is the following. Each module of cognition that is properly definable can be represented via the same underlying logic, which I have presented here in two slightly different 'incarnations'. The structures defined by this logic are models of the 'things' they represent, for instance visual objects. These models can be infinite, since they can potentially represent, for example, the infinity of objects we can recognize, events we can witness, and so on. The models defined by each module can be mapped onto a 'common' logical space, that of language: We can talk about what we see, smell, think, believe, etc.

This very informal discussion can be made more precise via the discussion of a well-known theorem of model-theoretic semantics, the *Löwenheim–Skolem theorem*. This theorem can be very roughly paraphrased in the following way: If a first-order logic has infinite models, then it is a corresponding countable infinite model. In our case, its import can be seen as follows. We may define several logical systems, each of them representing a single module of cognition. Each logic has the same underlying (and thus domain-general) syntactic and semantic principles. Each logic can define an infinite model: We may be able to recognize an infinity of (moving) objects, an infinity of sounds, realize an infinite of possible actions, and so on. Defined in this way, each logic/module appears to be an independent system, an internal model that potentially allows to represent how we can interact with the external world, but needs not to rely on 'external' inputs for these computations.

The Löwenheim–Skolem theorem tells us that even if we have an infinity of such logical systems, it is possible to define a more general logic which includes all of these modules in a 'common' logical space. More precisely, the *downward* part of the theorem tells us that, if a model is (countably) infinite, then this model may include an infinity of possible sub-models, themselves infinite. The *upward* part of the theorem tells us that for each (infinite) sub-model, we can find an extension of this model that includes the sub-model and some other elementary statements. So, if our 'main' model represents language, it will include models of other modules as proper sub-models (downward part); if a module such as vision can be represented via a model, then this model can be integrated inside the main model of language (upward part).

The conceptual import of this theorem can be dynamically interpreted as follows. We can assume that, for each (well-formed) visual computation, we can have a matching VRS in our model of vision. Each visual information unit can then be mapped onto the language model, and thus be part of a general model that includes other types of information (upward part). Conversely, for each linguistic unit so defined, a corresponding non-linguistic unit can be found, so that from the general model, we can move to the more specific model (downward part). This process can unravel over time: for each thing we see, we may have a corresponding noun, which we then associate to any object that has that shape, to put it in a very informal way. The same principle of closure can be defined for adpositions (and verbs): For each type of spatio-temporal relation between objects we 'see', we may have a corresponding adposition, which we associate to any relation that has that spatio-temporal structure, or 'shape'.

Both model (language) and sub-model (vision) will thus be expanded or updated over time, but the underlying (Boolean) structure representing these processes and their results will retain the same basic structure, as this update process will be guided by the same basic principles. Informally, these models can become quite 'rich' over time, but the basic *structural principles* by which their growth occurs remain the same, as a consequence of their recursive definition. In this regard, (full) recursion represents the possibility for the language to apparently expand *ad infinitum*, representing any type of information in a common space. Similarly, the relation between the language model and its sub-models, which takes the shape of interface relations/conditions, represents the possibility that language (recursively) emerges as a 'general' model, generated by the projection of all models of cognition into a 'neutral' logical space.[23]

I shall thus propose the following answer to the third research question:

A–3: *The nature of the vision–language interface is that of a bijection; recursive closure principles and interface conditions define what is unique to language.*

What distinguishes language from other modules of cognition is not the type of underlying structure, but two properties emerging from this structure and its ability to represent other structures in common space, (full) recursion and interface conditions. The answer I offered so far is virtually the same offered by Hauser *et al*. (2002), although the argument on which I have based my answer is relatively different, and perhaps places a greater emphasis on the *interaction* between recursion and interface conditions and their inherent 'logicality', as the kernel properties of FLN. This answer is also consistent with the considerations made by Pinker & Jackendoff (2005) and similarly-minded contributions to the FLN/FLB debate, since it suggests that language and other modules of cognition are quite more similar than it may appear at first glance.

The answer I offered so far might also offer an insight with respect to one important Biolinguistic problem, the emergence of language from an evolutionary perspective. I tentatively suggest the following way to look at this problem, assuming in advance that what I shall say in this paragraph is nothing more than a wild (and perhaps wrong) conjecture. If we take a logical perspective and compress the evolutionary millennia into a conceptual space, then the emergence of a FLN kernel, from an evolutionary perspective, occurs when the integration of different types of information into a common format emerges.

Pushing this speculative line to its logical limit, we might assume that at some point, roughly 200,000 years ago, our ancestors were (suddenly?) able to compare what they saw with what they said, and *vice versa*. From this initial step, which could be called the first step of induction, we might as well as assume that the $n+1^{\text{th}}$ subsequent steps followed suit over the next few hundred of thousand years, taking shape as the unraveling of the gamut of languages we can currently

23 This assumption leaves open the problem of 'how many' models make up our cognitive architecture that are integrated in this model. I leave open this question, but I assume that we can leave out the 'massive modularity' hypothesis typical of some evolutionary psychology literature. See Fodor (1998, 2000) for further discussion, however.

attest in the world (see e.g. Piattelli-Palmarini & Uriagereka 2005 for discussion and some references).

This single and yet very powerful emergent property could have arisen as the possibility (perhaps, necessity) to integrate different bits of information into an internally coherent (and perhaps optimal) representational/computational system. It might have been the case that language arose as the 'proof' that it is possible for different cognitive processes/modules to combine together into a unified, coherent cognitive architecture; thus it emerged entirely because of internal, structural pressures (Buzśaki 2006), although it became one tool (out of many) for humans to grasp and represent facts about the world, including the position of the things we see around us.

I shall leave these complex topics aside, and focus my attention back to our much more modest topic of discussion. Given the discussion I offered so far, I shall propose the following answer to the global research question:

Q-A: *The relation between spatial vision and spatial language is an isomorphism, as both models represent the same 'amount' of information via different types of information.*

This answer sums up the results of this section. Note that, while in this section I have suggested that language, broadly defined, describes a model which includes a model of vision as one of its proper parts, if we focus on spatial language, then this portion of language has the same structure and properties of vision; consequently, it (correctly) appears that vision and language are more similar than it seems, as observed in much literature. Much more could be said about this topic, as the discussion I have offered in this section can only be thought as a very preliminary attempt at refining a notion of FLN (and FLB, for that matter) and its emergence, from the point of view of 'space'. For the moment, though, I shall leave this discussion aside, and move to the conclusions.

5. Conclusions

In this paper, I have offered a novel proposal regarding the relation between vision and language with respect to 'space' — our understanding of things and their place in the world. I have argued that our spatial vision and language are quite abstract in nature, as they involve the processing of various types of information and their ability to individuate objects and the events they are involved in as well as the 'structural' relations that emerge from this process of individuation.

In doing so, I have offered a number of innovations on several closely related topics, including an updated review of the debate, a model-theoretic approach to vision which covers data usually ignored in the debate on 'space' (via the VRT proposal), a novel DRT treatment of adpositions, and a novel analysis of the vision–language interface, and what consequences this analysis has for a general theory of the language faculty.

The general picture I offered is one in which different models of cognitive processes can be formally defined in detail, and then embedded into a more general model of 'knowledge', modeled via a particular approach to Fodor's (1975) notion of 'language of thought', DRT, and the 'modularity of mind' hypothesis (Fodor 1983), although taken from a definitely more logical stance (as in e.g. Crain & Khlentzos 2008, 2009). Informally, language represents a 'neutral' logical space, a model of knowledge representation in which different concepts can be freely combined together, since they are already stripped of their 'external' constraints when they are represented in the corresponding models (e.g., Asher & Pustejovsky 2004, Asher 2011, and references therein). A similar reasoning holds for the articulatory-perceptual side of language. While we need to organize speech streams, say, into coherent units, the result of this process must then be organized into a coherent structure of syllables, words, and utterances which may be organized according to processes and relations not unlike those of other modules, and which are then mapped onto concepts, and thus lose their 'external' part. See Reiss (2007), Hale & Reiss (2008), and Samuels (2009) for discussion.

In this regard, language is the model that comes into being when all other 'sub-models' expressed by other modules of cognition are joined in a common logical space, and which might have emerged as the 'projection' of different cognitive modules into this common logical space. With respect to this neutral logical space, then, spatial language represents that fragment of space which represents spatial vision, i.e. our abstract representation of things in the world, whether this representation is veridical or not. As a consequence, the proposals I have made here, although still very preliminary in their nature, can be seen as offering a better picture not only on what is the nature of spatial representations in vision and language, but also on the logic behind the processes by which we combine together these representations, and what this tells us about the general architecture of mind and language.

References

Arsenijević, Boban. 2008. From spatial cognition to language. *Biolinguistics* 2, 23–45.

Asher, Nicholas. 2011. *Lexical Meaning in Context: A Web of Word*s. Cambridge: Cambridge University Press.

Asher, Nicholas & James Pustejovsky. 2004. Word meaning and commonsense metaphysics. Ms., University of Texas at Austin.

Barwise, Jon & John Etchemendy. 1990. Information, infons and inference. In Robin Cooper Kuniaki Mukai & John Perry (eds.), *Situation Semantics and Its Applications*, vol. 1, 33–78. Stanford, CA: CSLI.

Barwise, Jon & Jerry Seligman. 1997. *Information Flow: The Logic of Distributed Systems*. Cambridge: Cambridge University Press.

Beaton, Alan. 2004. *Dyslexia, Reading and the Brain: A Sourcebook of Psychological Biological Research*. Hove: Psychology Press.

Biederman, Irving. 1987. Recognition by components; a theory of human image understanding. *Psychological Review* 94, 115–147.
Bierwisch, Manfred. 1996. How space gets into language? In Bloom *et al.* (eds.), 1–30.
Blackburn, Paul, Johan van Benthem & Frank Wolter (eds.). 2006. *Handbook of Modal Logic*. Amsterdam: Elsevier.
Bloom, Paul. 2000. *How Children Learn the Meanings of Words*. Cambridge, MA: MIT Press.
Bloom, Paul, Mary A. Peterson, Lynn Nadel & Merrill Garrett (eds.). 1996. *Language and Space*. Cambridge, MA: MIT Press.
Bruce, Vicki, Patrick R. Green & Mark A. Georgeson. 2004. *Visual Perception: Physiology, Psychology and Ecology*, 4th edn. Hove: Psychology Press.
Burgess, Neil. 2006a. Spatial memory: How egocentric and allocentric combine. *Trends in Cognitive Science* 24, 603–606.
Burgess, Neil. 2006b. Spatial cognition and the brain. *Acts of the New York Academy of Science* 1124, 77–97.
Burgess Neil, Eleanor A. Maguire & John O'Keefe. 2002. The human hippocampus and spatial and episodic memory. *Neuron* 35, 625–641.
Burgess, Neil & John O'Keefe. 1996. Neuronal computations underlying the firing of place cells and their role in navigation. *Hippocampus* 6, 749–762.
Burgess, Neil & John O'Keefe. 2003. Neural representations in human spatial memory. *Trends in Cognitive Sciences* 7, 517–519.
Buzśaki, György. 2006. *Rhythms of the Brain*. Oxford: Oxford University Press.
Carey, Susan. 1992. Becoming a face expert. *Philosophical Transactions of the Royal Society of London* 335, 95–103.
Carey, Susan. 1994. Does learning a language require conceptual change? *Lingua* 92, 143–167.
Carey, Susan. 2001. Whorf versus continuity theorists: Bringing data to bear on the debate. In Melissa Bowerman & Stephen C. Levinson (eds.), *Language Acquisition and Conceptual Development*, 185–214. Cambridge: Cambridge University Press.
Carey, Susan & Fei Xu. 2001. Infants knowledge of objects: Beyond object files and object tracking. *Cognition* 80, 179–213.
Carlson, Laura A. 1999. Selecting a reference frame. *Spatial Cognition and Computation* 1, 365–379.
Carlson, Laura A., Terry Regier & Eric Covey. 2003. Defining spatial relations: Reconciling axis and vector representations. In van der Zee & Slack (eds.), 111–131.
Carlson, Laura A., Terry Regier, William Lopez & Bryce Corrigan. 2006. Attention unites form and function in spatial language. *Spatial Cognition and Computation* 6, 295–308.
Carlson, Laura A. & Emile van der Zee (eds.). 2005. *Functional Features in Language and Space: Insights from Perception, Categorization and Development*. Oxford: Oxford University Press.
Carlson-Radvansky, Laura A. & David E. Irwin. 1994. Reference frame activation during spatial term assignment. *Journal of Memory and Language* 33, 646–671.
Chierchia, Gennaro. 1995. *Dynamics of Meaning: Anaphora, Presupposition and the*

Theory of Grammar. Chicago: Chicago University Press.
Chierchia, Gennaro & Raymond Turner. 1988. Semantics and property theory. *Linguistics & Philosophy* 11, 191–216.
Coltheart, Max, Kathleen Rastle, Conrad Perry, Robyn Langdon, & Johannes Ziegler. 2001. DRC: A dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108, 204–256.
Coventry, Kenny R. & Simon C. Garrod. 2004. *Saying, Seeing and Acting: The Psychological Semantics of Spatial Prepositions*. Hove: Psychology Press.
Coventry, Kenny R. & Simon C. Garrod. 2005. Towards a classification of extra geometric influences on the comprehension of spatial prepositions. In Carlson & van der Zee (eds.), 149–162.
Coventry, Kenny R., Thora Tenbrink & John Bateman (eds.). 2009. *Spatial Language and Dialogue*. Oxford: Oxford University Press.
Cowan, Nelson. 1988. Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information processing system. *Psychological Bulletin* 104, 163–191.
Cowan, Nelson. 1995. *Attention and Memory: An Integrated Framework*. New York: Oxford University Press.
Cowan, Nelson. 2005. *Working Memory Capacity*. Hove: Psychology Press.
Craik, Kenneth. 1943. *The Nature of Explanation*. Cambridge: Cambridge University Press.
Crain, Stephen & Drew Khlentzos. 2008. Is logic innate? *Biolinguistics* 2, 24–56.
Crain, Stephen & Drew Khlentzos. 2009. The logic instinct. *Mind and Language* 25, 30–65.
Crain, Stephen & Mark Steedman. 1985. On not being led up the garden path: The use of context by the psychological parser. In David R. Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural Language Parsing: Psychological, Computational and Theoretical Perspectives,* 320–358. Cambridge: Cambridge University Press.
Cresswell, Maxwell J. 1985. *Structured Meanings*. Cambridge, MA: MIT Press.
Davidson, Donald. 1967. The logical form of action sentences. In Nicholas Rescher (ed.), *The Logic of Decision and Action*, 81–95. Pittsburgh, PA: University of Pittsburgh Press.
van der Does, Jaap M. & Michiel van Lambalgen. 2000. A logic of vision. *Linguistics & Philosophy* 23, 1–92.
van Eijck, Jan & Joost Zwarts. 2004. Formal concept analysis and prototypes. *Proceedings of the Workshop on the Potential of Cognitive Semantics for Ontologies*, 1–7.
van Eijck, Jan & Hans Kamp. 1997. Representing discourse in context. In Johan van Benthem & Alice ter Meulen (eds.), *Handbook of Logic and Language*, 179–237. Amsterdam: Elsevier.
Farah, Martha J. 2004. *Visual Agnosia*, 2nd edn. Cambridge, MA: MIT Press.
Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179–210.
Fodor, Jerry A. 1975. *The Language of Thought*. Cambridge, MA: Harvard University Press.
Fodor, Jerry A. 1983. *Modularity of Mind*. Cambridge, MA: The MIT Press.

Fodor, Jerry A. 1998. *Concepts: Where Cognitive Science Went Wrong*. New York: Oxford University Press.
Fodor, Jerry A. 2000. *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology.* Cambridge, MA: The MIT Press.
Fodor, Jerry A. 2003. *Hume Variations*. New York: Oxford University Press.
Fong, Vivienne. 1997. The order of things: What directional locatives denote. Stanford, CA: Stanford University Ph.D. dissertation.
Fujita, Koji. 2009. A prospect for evolutionary adequacy: Merge and the evolution and development of human language. *Biolinguistics* 3, 128–153.
Ganter, Bernhard & Rudolf Wille. 1998. *Formal Concept Analysis: Mathematical Foundations*. Berlin: Springer.
Gärdenfors, Peter. 2000. *Conceptual Spaces: The Geometry of Thought*. Cambridge, MA: MIT Press.
Geurts, Bart. 1999. *Presuppositions and Pronouns*. Oxford: Elsevier.
Gibson, James J. 1966. *The Senses Considered as Perceptual Systems*. Boston, MA: Houghton Mifflin.
Gibson, James J. 1979. *The Ecological Approach to Visual Perception*. Boston, MA: Houghton Mifflin.
Grätzer, George. 1978. *Universal Algebra*, 2nd edn. New York: Springer.
Grodzinsky, Yosef. 1990. *Theoretical Perspectives on Language Deficits*. Cambridge, MA: MIT Press.
Hale, Mark & Charles Reiss. 2008. *The Phonological Enterprise*. Oxford: Oxford University Press.
Hamm, Fritz & Michiel van Lambalgen. 2005. *The Proper Treatment of Events*. Cambridge, MA: MIT Press.
Hauser, Marc D., Noam Chomsky & William Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
Heim, Irene. 1982. The semantics of definite and indefinite noun phrases. Amherst, MA: University of Massachusetts Ph.D. Dissertation.
Herskovits, Annette. 1986. *Language and Spatial Cognition: An Interdisciplinary Study of the Prepositions in English*. Cambridge: Cambridge University Press.
Higginbotham, James. 2009. Two interfaces. In Massimo Piattelli-Palmarini, Juan Uriagereka & Pello Salaburu (eds.), *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*, 142–154. Oxford: Oxford University Press.
von Hofsten, Claes, Qi Feng & Elizabeth S. Spelke. 2000. Object representation and predictive action in infancy. *Developmental Science* 3, 193–205.
von Hofsten, Claes, Peter Vishton, Elizabeth S. Spelke, Qi Feng & Kerstin Rosander. 1998. Predictive action in infancy: Tracking and reaching for moving objects. *Cognition* 67, 255–285.
Hughes, Georges E. & Maxwell J. Cresswell. 1996. *A New Introduction to Modal Logic*. New York: Routledge.
Hummel, John E. & Irving Biederman. 1992. Dynamic binding in a neural network for shape recognition. *Psychological Review* 99, 480–517.
Hummel, John E. & Brian J. Stankiewicz. 1996. Categorical relations in shape perception. *Spatial Vision* 10, 201–236.

Hummel, John E. & Brian J. Stankiewicz. 1998. Two roles for attention in shape perception: A structural description model of visual scrutiny. *Visual Cognition* 5, 49–79.

Jackendoff, Ray. 1983. *Semantics and Cognition*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1990. *Semantic Structures*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 1991. Parts and boundaries. *Cognition* 41, 9–45.

Jackendoff, Ray 1997. *The Architecture of the Language Faculty*. Cambridge, MA: MIT Press.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. Oxford: Oxford University Press.

Jackendoff, Ray & Steven Pinker. 2005 The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, & Chomsky). *cognition* 97, 211–225.

Jackendoff, Ray & Fred Lerdahl. 2006. The human music capacity: What is it and what's special about it? *cognition* 100, 33–72.

Johnson-Laird, Philip N. 1983. *Mental Models*. Cambridge, MA: Harvard University Press.

Johnson-Laird, Philip N. 1992. *Human and Machine Thinking*. Hillsdale, NJ: Lawrence Erlbaum.

Kahneman, Daniel, Anne Treisman & Brian J. Gibbs. 1992. The reviewing of object files: Object-specific integration of information. *Cognitive Psychology* 24, 175–219.

Kamp, Hans. 1979a. Events, instants and temporal reference. In Rainer Bäuerle, Urs Egli & Arnim von Stechow (eds.), *Semantics from Different Points of View*, 376–417. Berlin: Springer.

Kamp, Hans. 1979b. Some remarks on the logic of change. Part I. In Christian Rohrer (ed.), *Time, Tense and Quantifiers*, 103–114. Tübingen: Niemeyer.

Kamp, Hans. 1981. A theory of truth and semantic representation. In Jeroen Groenendijk, Theo Janssen & Martin Stokhof (eds.), *Formal Methods in the Study of Language*, 277–322. Amsterdam: Mathematisch Centrum.

Kamp, Hans & Uwe Reyle. 1993. *From Discourse to Logic*. Dordrecht: Kluwer.

Kamp, Hans, Josef van Genabith & Uwe Reyle. 2005. Discourse Representation Theory. In Dov M. Gabbay & Franz Guenthner (eds.), *Handbook of Philosophical Logic*, 125–394. Dordrecht: Kluwer.

Keenan, Edward L. & Leonard M. Faltz. (1985). *Boolean Semantics for Natural Language*. Dordrecht: Reidel.

Keil, Frank C. 1989. *Concepts, Kinds and Cognitive Development*. Cambridge, MA: MIT Press.

Kim, In Kyeong & Elizabeth S. Spelke. 1992. Infants' sensitivity to effects of gravity on visible object motion. *Journal of Experimental Psychology: Human Perception and Performance* 18, 385–393.

Kim, In Kyeong & Elizabeth S. Spelke. 1999. Perception and understanding of effects of gravity and inertia on object motion. *Developmental Science* 2, 339–362.

Kracht, Marcus. 2002. On the semantics of locatives. *Linguistics & Philosophy* 25, 157–232.

Kratzer, Angelika. 1989. An investigation of the lumps of thought. *Linguistics &*

Philosophy 12, 607–653.
Kratzer, Angelika. 2007. Situations in natural language semantics. In Edward N. Zalta (ed.), *The Stanford Online Encyclopedia of Philosophy*, Fall 2011 edn. [http://plato.stanford.edu/entries/situations-semantics].
Kratzer, Angelika. To appear. *The Event Argument and the Semantics of Verbs*. Cambridge, MA: the MIT Press.
Krifka, Manfred. 1998. The origins of telicity. In Susan Rothstein (ed.), *Events and Grammar*, 197–235. Dordrecht: Kluwer.
Landau, Barbara. 1994. Object shape, object name, and object kind: Representation and development. *The Psychology of Learning and Motivation* 31, 253–304.
Landau, Barbara. 2002. Early experience and cognitive organization. In Lynn Nadel (ed.), *Encyclopedia of Cognitive Science*. Oxford: Wiley-Blackwell.
Landau, Barbara & James E. Hoffman. 2005 Parallels between spatial cognition and spatial language: Evidence from Williams syndrome. *Journal of Memory and Language* 53, 163–185.
Landau, Barbara & Ray Jackendoff. 1993. What and where in spatial language and spatial cognition. *Behavioral and Brain Sciences* 16, 217–265.
Landau, Barbara, Linda B. Smith & Susan Jones. 1992. Syntactic context and the shape bias in children's and adults' lexical learning. *Journal of Memory and Language* 31, 801–825.
Landau, Barbara, Linda B. Smith & Susan Jones. 1998. Object perception and object naming in early development. *Trends in Cognitive Sciences* 2, 19–24.
Landau, Barbara & Deanna S. Stecker. 1990. Objects and places: Geometric and syntactic representation in early lexical learning. *Cognitive Development* 5, 287–312.
Landman, Fred. 1991. *Structures for Semantics*. Dordrecht: Kluwer.
Landman, Fred. 2000. *Events and Plurality: The Jerusalem Lectures.* Dordrecht: Kluwer.
Landman, Fred. 2004. *Indefinites and the Type of Sets*. Oxford: Blackwell.
Levelt, Willem J.M. 1989. *Speaking: From Intention to Articulation*. Cambridge, MA: The MIT Press.
Levinson, Stephen C. 2003. *Space in Language and Cognition: Explorations in Cognitive Diversity*. Cambridge: Cambridge University Press.
Levinson, Stephen C. & Sergio Meira. 2003. 'Natural concepts' in the spatial topological domain — adpositional meanings in cross-linguistic perspective: An exercise in semantic typology. *Language* 79, 485–516.
Lewis, David. 1986. *On the Plurality of Worlds*. Malden, MA: Blackwell.
Link, Godehard. 1983. The logical analysis of plurals and mass terms: A lattice theoretical approach. In Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds.), *Meaning, Use, and Interpretation of Language*, 302–323. Berlin: Mouton de Gruyter.
Link, Godehard. 1998. *Algebraic Semantics in Language and Philosophy*. Stanford, CA: CSLI.
Maier, Emar. (2006). Belief in context: Towards a unified semantics of de re and de se attitude reports. Nijmegen: Radboud University Ph.D. dissertation.
Margolis, Eric & Stephen Laurence. 1999. Concepts and cognitive science. In Eric

Margolis & Stephen Laurence (eds.), *Concepts: Core Readings*, 3–82. Cambridge, MA: MIT Press.
Marr, David. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco, CA: Freeman and Co.
Marr, David & H. Keith Nishihara. 1978. Representation and recognition of the spatial organization of three dimensional structure. *Proceedings of the Royal Society of London B* 200, 269–294.
Mix, Kelly S., Linda B. Smith & Michael Gasser. 2010. *The Spatial Foundation of Language and Cognition*. Oxford: Oxford University Press.
Montague, Richard. 1973. The proper treatment of quantification in ordinary English. In Jaakko Hintikka, Julius Moravcsik & Patrick Suppes (eds.), *Approaches to Natural Language*, 221–242. Dordrecht: Reidel.
Munnich, Edward & Barbara Landau. 2003. The effect of spatial language on spatial representations: Setting some boundaries. In Dedre Gentner & Susan Goldin-Meadow (eds.), *Language in Mind: Advances in the Study of Language and Thought*, 113–155. Cambridge, MA: MIT Press.
Muskens, Reinhard. 1996. Combining Montague semantics and discourse representation. *Linguistics & Philosophy* 19, 143–186.
Nam, Senghou. 1995. The semantics of locative prepositional phrases in English. Los Angeles, CA: UCLA Ph.D. Dissertation.
O'Keefe, John. 1983. A review of the Hippocampus place cells. *Progress in Neurobiology* 13, 419–439.
O'Keefe, John. 1990. A computational theory of the hippocampal cognitive map. *Progress in Brain Research* 83, 301–312.
O'Keefe, John. 1991. The hippocampal cognitive map and navigational strategies. In Jacques Paillard (ed.), *Brain and Space*, 273–295. New York: Oxford University Press.
O'Keefe, John. 1996. The spatial prepositions in English, vector grammar, and the cognitive map theory. In Bloom *et al.* (eds.), 277–301.
O'Keefe, John. 2003. Vector grammar, places, and the functional role of the spatial prepositions in English. In van der Zee & Slack (eds.), 70–95.
O'Keefe, John & Neil Burgess. 1999. Theta activity, virtual navigation and the human hippocampus. *Trends in Cognitive Science* 3, 403–406.
O'Keefe, John & Lynn Nadel. 1978. *The Hippocampus as a Cognitive Map*. Oxford: Clarendon Press.
Parsons, Terence. 1990. *Events in the Semantics of English*. Cambridge, MA: MIT Press.
Phillips, Colin. 1996. *Order and Structure*. Cambridge, MA: MIT dissertation.
Piattelli-Palmarini Massimo & Juan Uriagereka. 2005. The evolution of the narrow language faculty: The skeptical view and a reasonable conjecture. *Lingue e Linguaggio* IV, 27–79.
Pinker, Stephen, & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95, 201–236.
Poggio, Tomaso & Shimon Edelman. 1990. A network that learns to recognize three dimensional objects. *Nature* 343, 263–266.
Pylyshyn, Zenon W. 1984. *Computation and Cognition: Towards a Foundation for*

Cognitive Science. Cambridge, MA: MIT Press.
Pylyshyn, Zenon W. 1989. The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition* 32, 65–97.
Pylyshyn, Zenon W. 1994. Primitive mechanisms of spatial attention. *Cognition* 50, 363–384.
Pylyshyn, Zenon W. 2001. Visual indexes, preconceptual objects, and situated vision. *Cognition* 80, 127–158.
Pylyshyn, Zenon. 2003. *Seeing and Visualizing: It's Not What You Think*. Cambridge, MA: MIT Press.
Pylyshyn, Zenon W. 2004. Some puzzling findings in multiple object tracking (MOT): I. Tracking without keeping track of object identities. *Visual Cognition* 11, 801–822.
Pylyshyn, Zenon W. 2006. Some puzzling findings in multiple object tracking (MOT): II. Inhibition of moving nontargets. *Visual Cognition* 14, 175–198.
Pylyshyn, Zenon W. & Vidal Annan Jr. 2006. Dynamics of target selection in multiple object tracking (MOT). *Spatial vision* 19, 485–504.
Quine, William Orman van. 1960. *Word and Object*. Cambridge, MA: MIT Press.
Ramchand, Gillian. (2008). *Verb Meaning and the Lexicon: A First Phase Syntax*. Oxford: Oxford University Press.
Regier, Terry & Laura A. Carlson. 2001. Grounding spatial language in perception: An empirical and computational investigation. *Journal of Experimental Psychology: General* 130, 273–298.
Regier, Terry & Mingyu Zheng. 2003. An attentional constraint on spatial meaning. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society*, 50.
Regier, Terry, Laura A. Carlson & Bryce Corrigan. 2005. Attention in Spatial language: Bridging geometry and function. In Laura A. Carlson & Emile van der Zee (eds.), *Functional Features in Language and Space: Insights from Perception Categorization and Development*, 191–204. Oxford: Oxford University Press.
Reiss, Charles. 2007. Modularity in the sound domain: Implications for the purview of universal grammar. In Gilliam Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 53–80. Oxford: Oxford University Press.
Reynolds, Jeremy R., Jeffrey M. Zacks & Todd S. Braver. 2007. A computational model of event segmentation from perceptual prediction. *Cognitive Science* 31, 613–643.
Riesenhuber, Maximilian & Tomaso Poggio. 1999a. Hierarchical models of object recognition in the Cortex. *Nature Neuroscience* 2, 1019–1025.
Riesenhuber, Maximilian & Tomaso Poggio. 1999b. A note on object class representation and categorical perception. *AI Memos 1679*, CBCL 183.
Riesenhuber, Maximilian & Tomaso Poggio. 2000. Models of object recognition. *Nature Neuroscience* 3 (suppl.), 1199–1204.
Riesenhuber, Maximilian & Tomaso Poggio. 2002. Neural mechanisms of object recognition. *Current Opinion in Neurobiology* 12, 162–168.
Rosch, Eleanor. 1975. Cognitive representations of semantic categories. *Journal of Experimental Psychology: General* 104, 192–233.
Samuels, Bridget. 2009. The structure of phonological theory. Cambridge, MA:

Harvard University dissertation.
van der Sandt, Rob A. 1988. *Context and Presupposition*. London: Croom Helm.
van der Sandt, Rob A. 1992. Presupposition projection as anaphora resolution. *Journal of Semantics* 9, 333–377.
Scholl, Brian J. 2001. Objects and attention: The state of the art. *Cognition* 80, 1–46.
Scholl, Brian J. 2007. Object persistence in philosophy and psychology. *Mind & Language* 22, 563–591.
Schwarzschild, Roger. 1996. *Pluralities*. Dordrecht: Kluwer.
Serre, Thomas, Lior Wolf & Tomaso Poggio. 2005. Object recognition with feature inspired by visual cortex. *Proceedings of 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, 994–1000.
Shutts, Kristin & Elizabeth Spelke. 2004. Straddling the perception–conception boundary. *Developmental Science* 7, 507–511.
Smith, Linda B., Susan S. Jones & Barbara Landau. 1996. Naming in young children: A dumb attentional mechanism? *Cognition* 60, 143–171.
Smith, Linda B., Susan S. Jones, Barbara Landau, Lisa Gershkoff-Stowe & Larissa Samuelson. 2002. Object name learning provides on-the-job training for attention. *Psychological Science* 13, 13–19.
Soja, Nancy N., Susan S. Carey & Elizabeth Spelke. 1992. Perception, ontology, and word meaning. *Cognition* 45, 101–107.
Soschen, Alona. 2008. On the nature of syntax. *Biolinguistics* 2, 196–224.
Speer, Nicole K., Kehna M. Swallow & Jeffrey M. Zacks. 2003. Activation of human motion processing areas during event perception. *Cognitive, Affective, & Behavioral Neuroscience* 3, 335–345.
Spelke, Elizabeth S. & Susan Hespos. 2001. Continuity, competence, and the object concept. In Emmanuel Dupoux (ed.), *Language, Brain, and Cognitive Development: Essays in Honor of Jacques Mehler*, 325–340. Cambridge, MA: MIT Press.
Spelke, Elizabeth S., Gary Katz, Susan E. Purcell, Sheryl Ehrlich & Karen Breinlinger. 1994. Early knowledge of object motion: Continuity and inertia. *Cognition* 51, 131–176.
Spelke, Elizabeth S. & Gretchen A. van de Walle. 1993. Perceiving and reasoning about objects: Insights from infants. In Naomi Eilen, Rosaleen McCarthy & Bill Brewer (eds.), *Spatial Representation*, 110–120. Oxford: Basil Blackwell.
Stalnaker, Robert. 1973. Presuppositions. *Journal of Philosophical Logic* 2, 447–457.
Stalnaker, Robert. 1999. *Context and Content: Essays on Intentionality in Speech and Thought*. Oxford: Oxford University Press.
Stankiewicz, Brian & John E. Hummel. 1996. MetriCat: A representation for subordinate and basic-level classification. *Proceedings of the 18th Annual Meeting of the Cognitive Science Society*, 254–259.
Stockhof, Martin, Jeroen Groenendijk & Frank Veltman. 1996. In Shalom Lappin (ed.), *Coreference and Modality: Handbook of Contemporary Semantic Theory*, 179–213. Malden, MA: Blackwell.
Svenonius, Peter. 2006. The emergence of axial parts. *Nordlyd* 33, 49–77.
Talmy, Leonard. 1978. Figure and ground in complex sentences. In Joseph H. Greenberg, Charles E. Ferguson & Edith A. Moravcsik (eds.), *Universals of Human Language*, vol. 4, 627–649. Stanford, CA: Stanford University Press.

Talmy, Leonard. 2000. *Towards a Cognitive Linguistics*. Cambridge, MA: MIT Press.
Trofimova, Maria. 2009. Case assignment by prepositions in Russian aphasia. Groningen: University of Groningen Ph.D. dissertation.
Tulving, Endel. 1972. Episodic and semantic memory. In Endel Tulving & Wayne Donaldson (eds.), *Organization of Memory*, 382–402. New York: Academic Press.
Tulving, Endel. 1983. *Elements of Episodic Memory.* Oxford: Clarendon Press.
Tulving, Endel. 2000a. Memory: Overview. In Alan E. Kazdin (ed.), *Encyclopedia of Psychology*, vol. 5, 161–162. New York: American Psychological Association and Oxford University Press.
Tulving, Endel. 2000b. Introduction to memory. In Michael S. Gazzaniga (ed.), *The New Cognitive Neurosciences*, 2nd edn., 727–732. Cambridge, MA: MIT Press.
Tulving, Endel. 2002. Episodic memory: From mind to brain. *Annual Review of Psychology* 53, 1–25.
Tversky, Barbara, Jeffrey M. Zacks & Bridgette Martin Hard. 2008. The structure of experience. In Thomas F. Shipley & Jeffrey M. Zacks (eds.), *Understanding Events*, 436–464. Oxford: Oxford University Press.
Ullman, Shimon. 1979. *The Interpretation of Visual Motion*. Cambridge, MA: MIT Press.
Ullman, Shimon. 1996. *High Level Vision*. Cambridge, MA: MIT Press.
Vaina, Lucia M. (ed.). 1990. *From the Retina to the Neocortex: Selected Papers of David Marr*. Boston, MA: Birkhauser.
Walle, Gretchen van der & Elizabeth S. Spelke. 1996. Spatiotemporal integration and object perception in infancy. *Child Development* 67, 621–2640.
Winter, Yoad. 2008. Between logic and common sense — Or: Dikkertje Dap meets Dr. Halfbaked. Talk presented at the Nijmegen Institute for Cognition and Information, Nijmegen (20 May 2008).
Wunderlich, Dieter. 1991. How do prepositional phrases fit into compositional syntax and semantics? *Linguistics* 29, 591–621.
Zacks, Jeffrey M. 2004. Using movement and intentions to understand simple events. *Cognitive Science* 28, 979–1008.
Zacks, Jeffrey M., Nicole K. Speer, Kehna M. Swallow, Todd S. Braver & Jeremy R. Reynolds. 2007. Event perception: A mind/brain perspective. *Psychological Bulletin* 133, 273–293.
Zacks, Jeffrey M. & Kehna M. Swallow. 2007. Event Segmentation. *Current Directions in Psychological Science* 16, 80–84.
Zacks, Jeffrey M. & Barbara Tversky. 2001. Event structure in perception and conception. *Psychological Bulletin* 127, 3–21.
Zacks, Jeffrey M., Barbara Tversky & Gowri Iyer. 2001. Perceiving, remembering and communicating structure in events. *Journal of Experimental Psychology: General* 130, 29–58.
van der Zee, Emile. 2000. Why we can talk about bulging barrels and spinning spirals: Curvature representation in the lexical interface. In Emile van der Zee & Urpo Nikanne (eds.), *Cognitive Interfaces: Constraints on Linking Cognitive Information*, 143–184. Oxford: Oxford University Press.

van der Zee, Emile & Jon Slack (eds.) 2003. *Representing Direction in Language and Space*. Oxford: Oxford University Press.

Zwarts, Joost. 2005. Prepositional aspect and the algebra of paths. *Linguistics & Philosophy* 28, 739–779.

Zwarts, Joost. 2010. A hierarchy of locations: Evidence from the encoding of direction in adpositions and cases. *Linguistics* 48, 983–1009.

Zwarts, Joost & Yoad Winter. 2000. Vector space semantics: A model-theoretic analysis of locative prepositions. *Journal of Logic, Language and Information* 9, 169–211.

Francesco-Alessio Ursini
Macquarie University
Macquarie Centre for Cognitive Sciences (MACCS)
Building C5C, Talavera Road
2113 North Ryde (Sydney), NSW
Australia
francescoalessio.ursini@students.mq.edu.au

★ REVIEWS ★

Quod Homines tot Sententiae — There Are as Many Opinions as There Are Men

Larson, Richard K., Viviane Déprez & Hiroko Yamakido (eds.). 2010. *The Evolution of Human Language*. Cambridge: Cambridge University Press.

by Lluís Barceló-Coblijn

"But as my conclusions have lately been much misrepresented, and it has been stated that I attribute the modification of species exclusively to natural selection, I may be permitted to remark that in the first edition of this work, and subsequently, I placed in a most conspicuous position — namely, at the close of the Introduction — the following words: "I am convinced that natural selection has been the main but not the exclusive means of modification."
This has been of no avail.
Great is the power of steady misrepresentation; but the history of science shows that fortunately this power does not long endure."
(Darwin 1870, final chapter of the sixth edition of *On the Origin of Species*)

Richard K. Larson, Viviane Déprez, and Hiroko Yamakido (Larson et al. 2010) have at last published one of the most eagerly awaited books on the evolution of human language, in which fourteen lectures have been collected from the *First Morris Symposium on Language and Communication* (held at Stony Brook University 14–15, 2005). The time elapsed between the conference and the publication of the volume is one of the reasons that make the book so interesting and long-awaited. The editors have chosen as their starting point the perhaps most controversial paper on language evolution of the last decade (Hauser *et al*. 2002), which could be secured as the volume's first chapter; very useful indeed, as it is cited and commented by most of the other contributors.

The Roman playwright Terence (Publius Terentius Afer, 195/185–159 BC) once said, "there are as many opinions as there are men".[1] And as soon as one reads the editors' introduction, one begins to feel that the variety of the arguments and points of view therein will be more than "several". Such feeling is indeed confirmed: The reader has in her hands fourteen different voices expressing different theories and presenting original arguments in order to support each of them — an attractive compendium to get an idea of the situation of current research on evolution of language.

I am grateful to Antoni Gomila for helpful comments. This work was supported by the HUM2007-64086 grant from the Ministerio de Ciencia e Innovación (Spain).

[1] The sentence of the title is thus his and it comes from the play *Phormio* (161 BC).

ISSN 1450–3417
Biolinguistics 5.3: 226–253, 2011
http://www.biolinguistics.eu

The first chapter, thus, is a reprint of Hauser *et al*. (2002, henceforth HFC) and will not be commented in depth in this review, since there already exist plenty of writing about it, including the intense debate consisting of Pinker & Jackendoff (2005), Fitch *et al*. (2005), and Jackendoff & Pinker (2005). In a nutshell, HCF provides an important framework for the study of language evolution. More specifically, HCF propose that "recursion and the mappings to interfaces" is a unique property of human cognition, constituting the only relevant aspect of the "faculty of language in the narrow sense" (FLN) — which, at the same time, is the core element of the "faculty of language in the broad sense" (FLB). This paper has provoked (and still does) many reactions in many different fields of study. It puts forward a provocative hypothesis about the human uniqueness — an issue that worry many scientists —, this time focused on the recursive capacity of human beings to produce limitless hierarchically structured sentences. According to their view, recursion is precisely the special element of human cognition and the element that non-human animals lack. What makes HCF special is that it represents an attempt to integrate in a single field of research — *biolinguistics* — empirical and theoretical issues that concern the biological study of language, the study of cognition in general, and its evolution.

The discussions that HCF has caused among linguists are well known, above all, due to the three different definitions of FLN the reader can find in the text. Whether it means "recursion only" or rather "recursion plus mappings" is something has given rise to a lot of opinions. Both Fitch but above all Hauser deal with this topic in their respective chapters and reveal that the original text "had to be cut to about half its original length" (p. 75), as an excuse for such central ambiguity.

Let me provide a sketch of the four-part structure of the volume:

(1) *Language architecture* (Chomsky, Jackendoff, Fitch, and Hauser);
(2) *Language and interface systems* (Gärdenfors & Osvath, Corballis, and Sperber & Origgi);
(3) *Biological and neurological foundations* (Dor & Jablonka, Piattelli-Palmarini, Lieberman, and Stromswold);
(4) *Anthropological context* (Tattersall, Bickerton, and Bingham).

The editors' purpose is to offer a storyline that provides some order within the chapters which deal with theoretical linguistics, genetics, biology, pragmatics, and so on, yet the borders of the frameworks of each contribution are not always clear, making evident the interdisciplinarity collected here. All of the contributors are reputed scientists, so it is clear from the beginning that this is not an introductory book, but a publication for advanced readers on these matters. In my view, this is both a weak spot (due to the possibility of losing some interesting details) and a virtue (because it forces the reader to take a look outside the traditional topics in linguistics and evolution).

After the editors' introduction and HCF, **Noam Chomsky**'s contribution is first. The text has been available on the internet until very recently for some time already, but the published version is more complete and carefully written.

Chomsky does not disappoint at all, in the sense that one can perceive that characteristic flavor of his style. This also means that a single reading is not enough to grasp it completely. In the beginning, Chomsky brings his point across when he makes clear that he is not in favor of an *adaptationist* view of language evolution. After a historical introduction about the birth of the term "biolinguistics", as in several of his papers, Chomsky invites the reader to consider whether language is the result of adding up "interfaces + recursion". This is important because one can easily follow the concept of language Chomsky has in mind. Unlike many texts about language evolution, Chomsky considers *language* human language only. What's more, he always talks about the linguistic system of *H. sapiens*. Hence, before this hominid there is no language but other kinds of communication systems. Even if the reader does not agree with that, it is of great appreciation that one has not to wait and make continuous suppositions until one finally understands what the author means by *language*. We will see throughout the book that this is unfortunately not a general rule.

Thus, according to Chomsky, it is worth considering the hypothesis that language is a computational system able to improve the cognitive capacity of human mind by means of the emergence of *unbound Merge* within the sensorimotor and the conceptual-intentional interfaces. In Chomsky's view, unbound Merge is a relatively new feature, evolved in modern humans only. Still in the minimalist framework, this mechanism of merging two elements into a new one again and again, is able to account for the structure language seems to show. The times when Universal Grammar was sophisticated and specified are gone. Now it contains minimal specifications to get the same results. The language does not matter because the underlying elements — Baker's (2002) *atoms* — are the same in Lakota, in Catalan, or in Basque. Here is where the reader can perfectly smell the Evo-Devo flavors: Inspired by Jacob's idea about genetics, Chomsky argues that it would be useful to adopt the basic concepts of evolutionary genetics in order to obtain a new picture of the events that affected language throughout its history. Thus, if minimal changes in control gene expression yield completely different biological forms, the same could be applied to language. In other words, linguistic variation would be just the result of minimal changes, being the underlying mechanisms the same in all languages and shared by all *H. sapiens*.

Another important idea, this time borrowed from Alan Turing, is the view of organisms as "living systems" that undergo the general laws of physics and chemistry, so that the possible forms are far from endless. According to this view, the superficial variety of organisms/languages is regulated by a *developmental genetic toolkit*.

This is a clear effort to incorporate some of the most remarkable ideas and theories from evolutionary biology to the studies about the evolution of the human faculty of language. According to Chomsky, there are (at least) three basic factors in language design: genetic endowment, external data, and principles not specific to language (Chomsky 2005). In this respect, I personally like Ott's (2007: 4) addition of a fourth factor that "concerns the embedding of the Language Faculty within the mind — that is, the way it interfaces with other components". It makes clear the biological frame in which language ontogenically develops, that is to say, *H. sapiens'* brain/mind and not any other.

All this is framed in an *internist* theory of language, a position one is almost forced to assume if one has — even if only generally — a formal conception of the mechanisms that structure language. The *Cartesian* stand Chomsky has taken traditionally is obviously defended here: In the beginning, "it was a language of thought". A student of evolutionary studies or comparative psychology would have immediately asked: If so, what about the rest of hominids? Could *H. heidelbergensis*, *H. neanderthalensis*, *H. erectus*, and so on, make use of that language of thought? The answer is "no", according to the final part of Chomsky's essay. At least *not* as *H. sapiens* does. Chomsky defends the idea that the conceptual system of our ancestors was different from that of non-human animals. But the inclusion of the rest of the members of the sub-tribe *Hominina* is almost never taken into account in the theories about language evolution. Chomsky adopts a skeptical view about a secondary and independent language of thought (p. 55), but, if our (above-mentioned) ancestors had a different, non-animal conceptual system, could they have had that kind of mental language? Bickerton is partly right when he notes (pp. 199–200, see below) that Chomsky and HCF almost never take into account the rest of the members of the genus *Homo* in their hypothesis (with the exception of some commentaries by Fitch and Hauser related to speech — but not language — and Neanderthals). That the rest of *Hominina* could vocalize (on their way) is almost certain, since all other primates can, and nothing on the fossil record indicates the opposite. So, in which place, as regards cognition, should we put those hominids? Another thing that is not completely clear is that this initial period for the language of thought could be misunderstood as a period of silence. In short, it is not clear at all whether there really was a moment zero for that silent language of thought only, or whether it was parallel to the vocal and/or gesture communication system those hominids could make use of.

Later on, Chomsky speculates and gives an example about a theoretic hominid called *Prometheus*[2] (p. 59) who, as the first member of his community endowed with unbound Merge, would have taken advantage of all its potential. We all, full-fledged modern humans, would be his descendents. Like *Prometheus*, we can make use of "duality of semantics, operator-variable constructions, *unpronounced* elements with substantial consequences for interpretation and thought" (p. 59; emphasis added). Here is maybe where the prose becomes a little bit messy. It is clear (it should be clear) that this is a metaphor; *Prometheus* was alone in his "internal linguistic" condition, so there was, at the beginning, no place for "unpronounced elements". What's more, it *was* a language of *thought*. Chomsky himself often cites Ian Tattersal, a paleoanthropologist who, among other interesting reflections, has argued that "the arrival of new behavioral or technological innovations has *not* tended to coincide with the appearance of new kinds of hominid. This actually makes considerable sense, for the only place in which a novelty can appear is *within* species" (Tattersall 2004: 22). In other words, it takes several generations, *within a species* for a novelty to be "discovered" and "exploited". It would not be necessary to state this (the fact that *Prometheus* did not

---

[2] Curiously, in Ancient Greek *Prometheus* means "forethought"; his mythological brother *Epimetheus* means "hindsight", literally "afterthought", but in the manner of a fool looking behind, while running forward.

exploit his potential language of thought), if we were not aware that sometimes we find comments on Chomsky's words because they have been taken literally. His example clearly does not help much to clarify his view. A radical reading of this passage could come to the conclusion that *Prometheus* produced unintelligible utterances to his own parents. This kind of literal reading can be found in this book in Bickerton's chapter (p. 202).

Next, we find **Ray Jackendoff**'s essay, which is, along with Lieberman's (see below), one of the most transparent in the presentation of the hypothesis he puts forward. Clearly, the author has a different theory of language and, hence, a different view of the evolution of this cognitive faculty. Jackendoff argues that a good strategy in order to explore the features of language and its evolution is *reverse engineering*. He first classifies the elements that compound language in four different departments: (1) things necessary for language, but that did not require genetic changes (e.g., lungs); (2) innovations in the human lineage useful for language or its acquisition that serves other general purposes (e.g., theory of mind); (3) aspects of language that are unique to humans, that are exclusively for language or its acquisition that required a change of the pre-existing primate structures (e.g., vocal tract; in this regard, the author agrees with Lieberman in that the vocal tract evolved *for* language); and (4) something altogether new and unprecedented in the primate lineage. The last one would be the right place for FLN, according to HCF (pp. 64–65).

Jackendoff underlines the fact that we need to have "analyses of other capacities to compare them to language". The problem is that there are no such analyses, just a few or largely abandoned ones (e.g., on music and on visual perception). In Jackendoff's opinion, other strategies like the comparative method advocated in HCF are insufficient. Departments (3) and (4) could be null, that is, "nothing *special* needed for evolution of language", though it is not his bet.

According to Jackendoff (p. 67), there are two kinds of theoretical architectures for language, syntactocentric (Chomsky's proposal) and parallel (Jackendoff 2002). The difference lies in the way both conceive the lexicon formation. For the former one, each item is an association of phonological, syntactic and semantic features, all of them embedded into a syntactic structure. So, syntax makes possible the connection of thought with vocalization. In this proposal, recursion is inserted between the interfaces. For the latter proposal, there are "independent principles of combinatoriality in phonology, syntax and semantics, each restricted to its proprietary structure" (p. 67).

The first and biggest problem Jackendoff sees in HCF's proposal is that "the whole generative syntactic system and the mappings to phonetic and logical form have to spring into existence more or less out of the blue" (p. 69). For HCF, recursion would be in department (4), whereas for Culicover & Jackendoff (2005), recursion is an element also of visual cognition, that is, belonging to department (3).

Jackendoff's proposal is original in the sense that the semantic/conceptual structure is the product of a combinatorial capacity, but at the same time independent of syntax. And here is where we find the vaguest part of his hypothesis: "[T]hought was highly structured in our ancestors [i.e. at least, the rest of the members of the genus *Homo*]" but "they couldn't express it" (p. 71). This kind of silent thought is not only the result of a combinatorial capacity, but its units are also liable to further non-syntactical combinations. According to the author, this is a preadaptation and its product, combinatorial thoughts, useful to be "shared". This is possible because the parallel architecture allows us to establish links between the phonological and semantic interfaces, without intervention of syntax. Hence, our ancestors would have a proto-lexicon (more or less à la Bickerton). As we will see in the review of Bickerton's chapter, the proto-language itself — in this case, the proto-lexicon — lends to the production of multiple vocalizations. Finally, syntax, "the capstone innovation", would have appeared in successive and gradual stages. The reader is referred to Jackendoff (2002) in order to learn more about the even more gradual stages the author proposes therein.

I think it is easy to grasp the great difference between Chomsky's view and Jackendoff's. Notwithstanding the final stage, the emergence of syntax is not described therein, which leaves the reader with a feeling of incompleteness. Back to the possibility of a kind of thought that is combinatorial, useful for sharing, but that couldn't be expressed at the beginning, the whole thing leads us to the next question, what did make possible to share this kind of inexpressible thoughts? Its usefulness for sharing?

The possibility is widely accepted that the vocal tract was prior to the modern capacity for language. However, most linguists are reluctant to concede a sophisticated vocal system to other hominids. Jackendoff concedes vocalization to hominids, in order to explain the emergence of linear order as a precursor of language (p. 71). But surely vocalization goes further back in time. As well as Chomsky, Jackendoff presupposes a surreptitious stage of silence where thought is already propositional but cannot be expressed. Looking at the rest of the primates, this seems an anomalous possibility. In any case, this is an interesting chapter that invites the reader to think about this plausible architecture of language.

In the next chapter, **W. Tecumseh Fitch** talks about something that is of great necessity in the debate originated by HCF: the meaning of recursion. Indeed, it was missing in HCF, since it was not published in a linguistics journal but in a journal for general science. The point is that although it could come as a surprise, recursion has three (really) different meanings, depending on the field of study — computer science, linguistics, or meta-mathematics. It is a fact that differences concerning the meaning of recursion arise immediately when linguists talk to mathematicians. They simply do not talk about the same thing. More or less the same can happen when one of the interlocutors comes from computer science (CS), though the differences may not be so strident. According to Fitch, recursive functions typically take "their own past output as their next input". On the one side, when defining recursion, as a term used in CS, we see that it "is one which calls itself" (the keyword is the verb *call*; p. 76); on the other side, in linguistics, recursion "has the property of self-embedding, that is, in which the same phrase type appears on both sides of a phrase structure rewrite rule" (p. 79). Clearly, here the keyword is *embedding*. This difference is crucial to understand why recursion is *different* in both fields, and in fact it is so crucial that a recent paper concerning the nature of recursion focuses precisely on this (Arsenijević & Hinzen 2010): It seems that recursion in linguistics necessarily implies embedding of elements. Finally, in meta-mathematics, there is a long tradition in the study of recursive functions, something that in some cases could surprise non-mathematicians, since, as Fitch noted, there are some iterative functions, and even non-recursive functions, which are included in the set of recursive functions. Fitch notes that mathematicians are more concerned with computability and not with whether a function recalls itself or implies embedding. These seem to be simply very different things and in fact a new label, *computability theory*, replaces the old terminology.

Another factor related to these definitions is whether or not there are tree representations behind the structures or outputs resulting from this operation. In the CS case, Fitch argues the answer is "no", since nowhere in the software or the code is there any implicit tree diagram. This might be something people just draw as an aid. Hence, the tree is not in the code. However, this is not true for linguistic theory, where the tree diagram is important and explicit. This remark helps Fitch to bridge another famous debate on whether tamarins and starlings have recursion (or lack it), arisen by virtue of the results published in Fitch & Hauser (2004) in which they put to the test cotton-top tamarins in order to see whether or not they are capable to process different kinds of grammars — a Finite State Grammar (FSG) and a Phrase State Grammar (PSG). This issue is brought up again by Hauser in this volume, too (p. 97).

The most extended interpretation (mine included, I confess) was that it was a test for recursion. What's more, it could have even been interpreted somehow as a kind of experimental proof of the *pumping lemma,* so many different readings were possible, depending on the reader's background. Anyway, immediately a great debate arose (e.g., Kochanski 2004) and other scholars put to the test humans (Perruchet & Rey 2005, who found that humans are not that good at learning a PSG) and starlings (Gentner *et al.* 2006, concluded that starlings can process a context-free grammar; but van Heijningen *et al*. 2009 disagree). Contrary to these interpretations, Fitch argues that Fitch & Hauser (2004) was not a test for recursion, that the word 'recursion' was even not mentioned in the paper. Fitch shows quite convincingly that both A^nB^n (PSG) and $(AB)^n$ (FSG) grammars "can be represented recursively". Nonetheless, it is also true that the picture of the grammars in Fitch & Hauser (2004: 378) clearly shows a classical diagram of a center-embedded grammar, so that the risk of misinterpretation was more than high.

What I find particularly interesting and of great value in Fitch's contribution is his aim to put some order within the terminology of the field, and at the same time his effort to build bridges to other fields of science. Abstract concepts like *recursion* are sometimes the seed of sterile debates simply because different people have different conceptions of the same term. I'm sure that researchers who work every day in interdisciplinary labs will appreciate this kind of work.

Next, in the fourth chapter, we find **Marc Hauser**'s contribution. The text takes a personal tone and, right from the beginning, the author advises the reader that these are his opinions (and not HCF's). First of all, Hauser defends his work done on animal communication, in response to Bickerton's (2007) "puzzling point" minimizing the work Hauser, Fitch, and Chomsky have done in this particular field (pp. 92–93). Next, the author rejects the interpretations of HCF as a paper that "flats out animal communication", recalling that he is still working in animal communication (p. 93). And so there are three more sections defending that: (i) the sensory-motor system as a homologue or analogue with other animals' system is a hypothesis; (ii) HCF's hypothesis is *not* recursion only; (iii) he is not a closet minimalist. By the way, an interesting last comment is that there exists a last chapter of this saga on the web: The reader can find another paper — this time the authors' order is "Chomsky, Hauser and Fitch (2005)" (p. 95) — where a response to the second part of Pinker & Jackendoff (2005), which focus on the Minimalist Program is provided. I'm sure that the reader interested in such debate will appreciate this *last* release — it is five years old, yet written at the time these communications were made.

Finally, Hauser deals, once again, with the distinction between FLN and FLB, what it is useful for, and why it should be taken into account. The author defends the usefulness of this strategy of putting the elements into one set or into another. The status of each element, says the author, has not to be permanent: As experimental research offers further empirical evidence, an element of FLN could be moved into FLB. Let us remind the reader that Fitch *et al*. (2005: 181) noted that FLN "could possibly be empty if empirical findings showed that none of the mechanisms involved are uniquely human or unique to language". The last part is a reflection about the above-mentioned experiment with cotton-top tamarins (Fitch & Hauser 2004). Hauser recognizes that such a grammar was not the best choice, since many other mechanisms "could underlie this competence" (e.g., the mechanism of counting argued for by van Heijningen *et al*. 2006).

Although here we will find neither a new theory nor a new hypothesis, the reader will find some answers to those questions that arise in reading the cited papers. It may not be the most spectacular essay of the volume, but these new pieces of information about Hauser's intentions and posterior reflections provide the reader with a human perspective that is missing so often in the scientific literature.

In the second part of the book, **Peter Gärdenfors & Mathias Osvath** make a contribution concerned with the evolutionary stage of the hominid mind, when there was yet no language (neither oral nor mental). The authors talk about a time prior to the emergence of symbolic thinking. They agree that *H. sapiens* is the only animal whose use of symbolic language has been proven. Thus, Gärdenfors & Osvath aim to deepen our knowledge of the forces behind language evolution. Their hypothesis focuses on the Oldowan culture (for that matter, see also my comments below on Bickerton's contribution) for the first stages of which, as they clearly state, there is no recognized author. This constitutes a problem because at that time there were many species of hominids. The authors follow Plummer (2004) on that matter, but the reader has to be aware that not only *H. habilis* and *H. erectus* could have made that lithic industry, but also *Parathropus*[3] according to

[3] It comprises at least three species: *P. aethiopicus*, *P. boisei* and *P. robustus*. This genus is be-

an examination of the hand of parathropines (Susman 1988). This is important because Gärdenfors & Osvath's hypothesis turns on "the Oldowan culture", unaware that this could affect both genera *Homo* and *Paranthropus* with very different results.[4] Thus far, the Oldowan culture becomes the ecological niche for some of these hominids, which could act as one of the driving forces towards symbolic thinking. Nevertheless note that, for Gärdenfors & Osvath's hypothesis to be viable, one has to assume that the same *force* has very different consequences on genetically very close co-existing species (a point, thus, indirectly in favor of evo-devo theses).

Their hypothesis is as follows: *Prospective cognition* precedes symbolic thought and is based in two kinds of thoughts, *cued* mental representations (CMR) and *detached* mental representations (DMR). The former refer to present objects, whereas the latter refer to non-present objects or events. DMR could be, according to the authors, one of the novelties of the frontal lobes, since these parts of the brain have been linked to activities like planning and fantasizing. Again, arguments such as these have to be taken very carefully. The current role of the frontal lobes could differ from their role in those days.[5] DMR are related to Hockett's *displacement*, though slightly different (p. 105), and imply the existence of an *inner world* (the collection of detached mental representations). DMR seem to be the basis of the ability "to envision various actions" which, according to the authors, is a requirement for *planning* (p. 106).

Gärdenfors & Osvath note in passing that even chimpanzees show the ability for planning when preparing tools for fishing termites. This would be a case of *immediate planning*, whereas *prospective planning* must have a detached representation of *future needs*. Although the authors do mention *H. habilis*, they bet indeed for *H. erectus* as the hominin showing such mental abilities (always following Plummer 2004). According to the authors, this hominin would have been able to carry lithic tools, to divide the labor within the members of the group according to their aptitudes, and to hunt and gather Although Gärdenfors & Osvath do not mention this, it's highly likely that *H. erectus* hunted micro-fauna. In any case, their view is that prospective cognition was a necessary evolutionary novelty.

This is the part of the paper I find more well-grounded and fitting best with paleoanthropological data. However, the linking to *H. sapiens*'s language is still remote. The way in which Gärdenfors & Osvath build the bridge is by appealing to the notion of *cooperation* as the element favored by language (p. 111): Symbolic language favors cooperation about future goals. I have no problems accepting that *prospective cognition* could have been an important element in the evolution of modern human cognition, but to resort to abilities like "manipulation of attention" (proposed by Tomasello, 1999: 131) or "sharing visions" does not

lieved to be a parallel line to the genus *Homo,* which reached an evolutionary cul-de-sac and died out. Other scholars believe they all three are should be included within the australopithecines.

4 Additionally, recent findings support the suspicions that the genus *Australopithecus* already made and use tools for scavenging 3.39 million years ago (McPherron *et al*. 2010).

5 In this line of argumentation, it has been argued that *H. floresiensis* could have had some kind of sophisticated mental abilities (Falk *et al*. 2005). See below, on Bickerton's essay, for similar observations.

explain completely how it is possible that language has the syntactic, phonological, or semantic structure it has. Note that in this context, *symbolic communication* is still not "virtually equivalent to language", as Tattersall says. Thus, the gap from *H. erectus*'s way of communication to current modern language, understood as our cognitive faculty, is too broad to be covered simply by arguments on the usefulness of sharing symbols through communication.

The seventh chapter is **Michael Corballis**' paper. Corballis tries to reconcile Chomsky's and Bickerton's theories (as we will see, Bickerton explicitly says these are simply incompatible). As many authors in the book, Corballis takes as his starting point the extant hominins of 2 million years ago (mya), explicitly *H. rudolfensis* and *H. habilis*,[6] and the lithic culture of that time, that is, the Oldowan culture. Another trait he considers important is bipedalism, something that characterizes the genus *Homo.* Let me update a bit on this point: Recent work on that matter claims that bipedalism is not an innovation of the genus *Homo*, since a previous ancestor, the *Ardipithecus* (4.4 mya), was already biped (see especially Lovejoy (2009) and Lovejoy *et al*. (2009a, b). Additionally, *knuckle walking* is different in gorillas and in chimpanzees, a fact that suggests that the typical locomotion of these two great apes are (independently appeared) derived traits of these species rather than an ancestral trait, thus suggesting a common bipedal locomotion in our distant ancestors.

Corballis thinks that both Chomsky's and Bickerton's theories can be reconciled "if it is supposed that language itself evolved gradually, but it was based in the first instance on manual gestures, with gradually increasing vocal involvement" (p. 115). I do not see clearly how this can fix the problem, since the point is that both Chomsky and Bickerton basically agree about the fact that vocalization and its physical apparatus were already part of our ancestors, before modern language was a reality. So, it is not a problem of the modality of the output, but a different vision of the way the computational mechanism underlying syntax evolved.

It is said that the classics do not fail; and, Corballis, as other authors in this book, resorts to some classic paleoanthropologic scenarios in which (i) savannah replaces dense forests; (ii) there is an increase of brain size, "driven by selection for such cognitive abilities"; and (iii) a protolanguage (in Bickerton's sense) increases its sophistication until it reaches the current state (p. 116).

As in any other adaptationist hypothesis, this process is gradual and takes place always through natural selection. Corballis argues that this is "reasonable", but he does not explain why. In fact, the author appeals to a famous Chomsky quote: "It would be a serious error to suppose that *all* properties, or interesting structures that evolved, can be "explained" in terms of natural selection" (Chomsky 1975: 59, p. 117 of the present volume; emphasis added). This and other commentaries have been interpreted as suggesting that Chomsky is against, or that he rejects, natural selection as a driving force in evolution (e.g., Johansson 2005: 161). The emphasized word *all* in that sentence, in that context, clearly points out that some properties can be explained by natural selection, while

[6] There is controversy about whether or not both hominids belong to the same species.

others cannot. In any case, Chomsky is not original in this way of thinking, as we can see in the introduction of this review, since Darwin himself would have agreed.

Corballis assumes with Bickerton and Chomsky that modern language (in the author's sense, but, a necessarily *orally* externalized language) appeared "with or even after *H. sapiens*" (p. 116). Such "after" is quite interesting, since one may wonder how, under normal conditions, it is possible that any human of any land can acquire any human language if this capacity did not arise from the beginning in the same African population group. According to Corballis, the answer is straightforward: "[I]t was not language itself that emerge with *H. sapiens*, but rather the capacity for autonomous speech" (pp. 115–116). This would be possible since, according to Corballis' hypothesis, syntax would have appeared in a gradual process while, the expression channel for the output was manual, rather than vocal.

One has to acknowledge the continuous effort Corballis makes in trying to integrate his ideas and theories to current paleoanthropological data, and this can be easily detected throughout his work. It is a difficult task, since the field continuously evolves as new findings are published in many fields of study. However, sometimes the author takes as evidence for his hypothesis some data — for example, the *hypoglossal canal* — which Corballis knows to be controversial (p. 117). It is even more surprising when the explanatory power of this physical trait has been put into question by several scholars specialized in speech evolution (e.g., Lieberman 1999, Fitch 2000). For Corballis, it is reasonable to assume that modern speech mechanisms were "incomplete" in Neandertals and the common ancestor they have with modern humans (p. 117). The author argues that a piece of evidence in this direction is the human *FOXP2* gene, which he views as a novelty of the species *H. sapiens*.[7]

Corballis' own gestural hypothesis is largely grounded in the discovery of mirror neurons, which fire in both hand and mouth movements, and in the recognition of these movements in conspecifics. Such neurons have been detected in the F5 area of monkeys' brains, but still not in humans — though there is a lot of indirect data suggesting their presence in our brains. The author notes that grasping movements even "affect the kinematics of speech itself" (p. 120), which is taken in support for the motor theory of speech perception (Liberman *et al*. 1967). According to this theory, speech sounds "are perceived in terms of how they are produced". It goes without saying that this theory is quite controversial, though it has, in my view, some good points.

Up to here, the evolutionary theoretical background is quite robust, in general lines. Further, Corballis speculates about a possible scenario, always taking into account the *mirror system* as the basis for the further development of language. Thus, according to him, communication was basically gestural. In Corballis' hypothesis, bipedalism is a crucial element for the freedom of hands. The author further speculates that, as the technology of tools develops, language and tool-making are in conflict, due to the fact that both activities require the use of

[7] A fact refuted in the last publication on this matter (Burbano *et al*. 2010). Some commentaries around this gene are made below, on Lieberman's contribution.

hands. Such an adaptationist story, however, is forced to exaggerate the role of an activity like tool-making and the time those hominids (all members of the crowd) should invest during millions of years, in order to make of a cultural activity a driving force, to such an extent that it finally acts on the genome and its subsequent development. Another weak point is that, even if this was really so, there is no reason why no other hominid followed such path. When *H. sapiens* left Africa, an encounter with *H. neanderthalensis* at the region of Kebara took place. Both cohabited the region during thousands of years; they are believed to have been in contact, that they possibly had trading relationships, and now we know that they were able of eventually interbreeding (Green *et al*. 2010). But, even before we knew this last incredible piece of genetic data, the archaeological and fossil information already suggested that the two hominids were not that different.[8] Instead, Corballis contends that "the final conversion to autonomous speech may have been an invention (Corballis 2002) or, as suggested above, it may have resulted from the *FOXP2* mutation (Corballis 2004)" (p. 123). Again, we find reduction to a single factor and overlooking other species within the same context.

The next chapter is written by **Dan Sperber & Gloria Origgi**, and it covers an aspect of language which is quite interesting: pragmatics. The authors show that contextual factors play an important role in the way we interpret an utterance, and how it is possible that even sharing the same code does not guarantees that we all process that utterance likewise. The reader not familiar with Sperber's *Relevance Theory* (e.g. Sperber & Wilson 1986) should know that it largely builds on the Gricean philosophical theory of language. Sperber & Origgi confront two models: the code model, based on the fact that sentences are sound and meaning pairs, and the inferential model, which states that the inferential information we get from the context is relevant for our final representation of the utterance.

Although, according to the authors, both models agree that languages are codes with a recursive grammar (p. 125), the inference model includes explicitly what the authors call *naïve psychology*, which includes the ability to attribute mental states to others. This is because humans seem to "spontaneously interpret one another's behavior […] as belief-guided fulfillment" (p. 126). Sperber & Origgi point out the importance of the continuous inferences we make in our communication acts, how they are intervened by the context, how communication can fail if the communicator cannot fulfill her intention "by making it manifest to the hearer" (p. 126) — "or to the beholder", in the case of sign languages, we could add.

The authors acknowledge that the manipulation of mental states can be useful, but they observe that this mechanism is "cumbersome"; instead, overt communication, where both actors (communicator and addressee; it is interesting the use of *communicator* instead of the classic *emissor*, maybe because one can be *emissor* without voluntarily being *communicator*[9]) "are intent on comprehension"

---

[8] Remember that both hominins share 99.5% of the genome.

[9] But see Seyfarth & Cheney (2003: 147) for a different notion of communication, where even unintended acts are taken as active parts of communication (more in tune with Claude Shannon's mathematical theory of communication): "Although the frog has no goal of

and hence, the transmission of information becomes successful at low cost. An interesting aspect of the authors' proposal is that in their model, "a fragmentary coding is sufficient", contrary to the code model which has to encode the information unambiguously.

The authors defend the idea that humans have an inferential model and animals do not. Animal codes would be closer to the code model, since both communicator and receiver must share the code. Any difference would lead to potential errors. Recall, for example, the alarm calls in vervet monkeys. According to Sperber & Origgi, this is an example of a genetically transmitted code. It is their opinion that such codes are counter-adaptive (p. 127). I do not agree with that, since this statement is made from a strong anthropocentric point of view. Alarm call systems cannot be counter-adaptive, since so many different extant species have this kind of communication system. That our system looks much better to us, to our human logic, is a different issue. If extant species do have an alarm call system, it is because it has been beneficial.

Anyway, Sperber & Origgi observe an important difference between genetic systems and inferential systems: The former does not easily allow the incorporation of new elements, whereas the latter "does not require that the communicator and audience have the same semantic representation of the utterance" (p. 128). What's more, an "ad hoc meaning is contextually constructed" (p. 128). I think this is an important observation in order to differentiate some well-known animal communication systems from the human communication system. This fact increases the sophistication of the system; however, as the authors observe, it does not "protect" the users from potential misunderstandings (p. 129).

Finally, Sperber & Origgi propose an imaginary situation in which the communicator has a more sophisticated system than the receiver, and it seems that communication does not fail. Here, the communicator could represent the first generation endowed with a syntactic device (and this reminds us of Chomsky's *Prometheus*), a device which allows the holder to go beyond the coding possibilities of the hearer. This does not represent a problem for them to communicate, while the contrary would not be true. Next, the authors affirm that the holder of such a new device and "her co-mutants [i.e. subsequent generations] communicate more effectively than other members of their community" (p. 130). Again, in my view, the authors confound the fact of having devices more complex syntactically and semantically, with the fact of being better at communication. Complexity does not always mean better results, especially in the light of ecology. The communicative systems of our ancestor and related species of hominids worked effectively enough for their communicative purpose, and the proof is that those hominins could occupy an ecological niche for thousands of years. In my opinion, their systems were not that bad, at least not for their communicative necessities. They were different, possibly qualitatively different, but good enough for the recipient, their minds. Apart from anything else, as Tattersall (2004) notes, the

---

communicating to the bat, communication occurs nonetheless, as bats take advantage of a lucky accident and extract useful information from a signal that evolved for entirely different reasons".

place where a novelty can appear is *within* species and usually it has to pass a very long period of time until the species discover its potential, so that, when they can exploit it, a considerable group of members of the community — if not a majority of them — already share such novelty.

In any case, Sperber & Origgi's contribution shows the power of inferential systems when working along with a linguistic system. It also informs us that this is an important aspect that should be covered by both theoretical and empirical biolinguistic research. I think the reader interested in communication and its evolution will appreciate this line of work.

The next chapter is **Daniel Dor & Eva Jablonka**'s contribution. The authors present an original hypothesis, quite different from what's proposed in the rest of the book. It has a well-grounded background in genetics, and it is framed in the light of evo-devo, though they pay more attention, in fact, to the development of the phenotype. Dor & Jablonka use a special notion of language: According to them, language is a collective *invention*, which "culturally evolved before its speakers were specifically prepared for it on the genetic level" (p. 136). In their particular notion, language is something that, in its last stage, "was already out there, as an object for learning" (p. 146).

Dor & Jablonka argue that this has been possible because "the social world evolved to the point that collective inventions became possible" (p. 136). Although the authors do not make explicit what would be required to properly speak of "invention", they insist that there are inventors of language, and that not everybody can become an inventor — the truth is that the level of abstraction — required to grasp the notion which is behind the "invention" — is sometimes not so clear. Thus, the key in their hypothesis is the genetic and neuronal *plasticity* of the human condition. This term refers to the "ability of a single genotype to generate, in response to different environmental circumstances, variable forms of morphology, physiology, and/or behavior" (p. 137). It seems that in every species there are individuals who have more plasticity than others, and this factor gives them the possibility to adapt to changes in the environment. The authors give the example of Kanzi, the bonobo, who grew up within special conditions so that pre-existing components of his developmental systems were reorganized (p. 137). The authors note that this is a complex process in which other mechanisms and elements play particular roles:; for example, the *attractors* are "stabilizing end-states towards which the system seems to "strive"" (p. 137). Further, *canalization* consists in the adjustment of developmental pathways by natural selection and, as the authors note, the opposite to *plasticity* to some extent.

In this contribution, Dor & Jablonka offer a new scenario for the evolution of language. The truth is that it is highly speculative; the authors make a lot of assumptions, almost without citing where their conclusions come from. Thus, we have to assume that their thesis is based on the notions just mentioned, *plasticity* in particular. It is the authors' opinion that the members of the community are inventors. An important point is that they always talk about humans, and never about other hominins, so that we do not exactly know whether the term "human" is mentioned abstractly or whether the entire language evolution took place in modern humans only, as it would be if we read the text literally. In any case, Dor

& Jablonka start from the point that humans try to solve problems and therefore invent new words. Interestingly, some "problems emerged as systemic consequences of the development of language" (p. 140). On the other hand, "the community gradually sophisticated its world-view adding new linguistic categories". *How* is something not explained. It seems that those categories were invented by an inventor and learned by the hearers. Following this process, "language developed into a system of rules" (p. 141). Again, we have to suppose the system was invented thanks to plasticity. A rule system leads to a major stability, and this, at the same time, led to an increase of plasticity.

Suddenly, the authors change the topic, and they talk about the evolution of languages, the emergence of slang and jargon in linguistic communities and how this helps to social secrecy, something that makes hazy the concept of *language* they have been talking about. When the authors refer to linguistic changes caused by phenotypic variations, they choose examples from phonology or speech, never from syntax. This is possible because it seems that the rules, in their scenario, simply were invented and then suffered a process of "social negotiation and struggle" (p. 142).

The reader will find that the notion of language is quite different, it has to be *learnable*, but not everybody can learn it, since there is variation in plasticity. However, people who could not learn the more complex system could perhaps learn part of it, at least to reach some level of comprehension (p. 145).

Reading this chapter, neither the temporal frame their hypothesis covers nor the species to which it applies are ever clear enough: Sometimes it seems that humans are *H. sapiens*, but sometimes this is doubtful. Moreover, I missed some bibliography to ground their many assumptions and speculations in the second part of the paper.

In the tenth chapter, the reader will find **Massimo Piattelli-Palmarini**, a veteran in this field. In this contribution, he summarizes the history of science (especially biology) in relation to linguistics, to show how linguistics has changed the way languages have traditionally been considered and observed: from collections of treasures to natural objects. After each part of his discourse, Piattelli-Palmarini draws a conclusion in the form of a lesson. So, when he talks about the first steps taken by the people behind *string theory*, the lesson is that linguists have to encourage empirical research and pursuit of new ideas, even when some of these ideas could have at first sight a "dimly" conceptual content; or that we should not put any limit to the level of abstraction, if such is required. Piattelli-Palmarini revises the parallels between linguistics and other scientific fields. Thus, the reader will be reminded that at the beginning there were languages and philology, until the notions of I-language and E-language appeared (still controversial, by the way). The former paid attention to the tacit knowledge of language and this changed the study language evolution, since the object of study was now a cognitive capacity rather than a prescriptive, to some extent artificial, grammar. The author guides the reader through the ages in which some linguists decided to pay attention to other disciplines in order to get new ideas, which could help them explain the structures underlying natural languages. It was then, tells us Piattelli-Palmarini, that syntax took a central position: first the generative grammars, and then the minimalist program. The author then focuses on *edge features* (Chomsky 2008) and the operations they carry out, stressing the importance of this concept as well as the notion of *phase* ("self-contained derivational domains, characteristically nested one into the other, that are simultaneously sent to the two [sensori-motor and conceptual-intentional] interfaces"; p. 151).

Piattelli-Palmarini agrees about that recursion is an essential element of human cognition, and especially of language. He terms "the age of specificity", when generative grammar, the modularity of mind, visual cognition, and the Chomsky hierarchy were established. He also observes that "[p]ossibly the right formal characterization still eludes us, or possibly there cannot be any such purely formal characterization, because of inherent bio-evolutionary contingencies" (p. 153), or in other words, *principles not specific to language* (also known as the third factor). Piattelli-Palmarini does not believe either in gradualist or functional explanations of language evolution (or organisms, for that matters); as a biologist, he rather contemplates "the biological picture" as quite complex, "multi-faceted", and therefore he believes that biolinguistics must incorporate new ideas and models (p. 157).

In conclusion, Piattelli-Palmarini has summarized the essentials of the last forty years of research on language evolution. The notions and conceptions are quite clear, and so are the goals: the understanding of the biological principles and structures underlying the cognitive faculty of language. If this supposes to change the whole traditional paradigm, so be it.

Philip Lieberman is also one of the veteran authors in language evolution. Quite impressively, Lieberman defends the same hypothesis after forty years, concerning the possibilities of the Neandertal vocal tract (Lieberman & Crelin 1971). In Lieberman (2002) we find an original hypothesis, built onto the knowledge accumulated in these years. Lieberman's approach focuses on *basal ganglia,* subcortical structures that, far from being old or static during the time, have evolved in a particular way in humans. In passing, Lieberman proposes a new term, *reiteration*, which subsumes the properties of *recursion* à la HCF, but it "is expressed outside the domain of language when we change the direction of a thought process as well as in seemingly unrelated activities such as dancing" (pp. 163–164). Besides the interest of the idea — no doubt important in the debate concerning the limits of the range of action of *recursion* in human cognition —, the question is whether or not such a new theoretic term is indeed necessary, given the close resemblance with *recursion*: both entail nested hierarchical structures (p. 164). One difference, mentioned above, is that iteration works outside language. A second difference, according Lieberman, is that "iteration instead generates the sentences and semi-sentences that can be observed in real life by inserting relative clauses, […] and other elements without the torturous and often arbitrary operations of traditional generative theories" (p. 164). The operations which entail *reiteration*, following Lieberman, are thus linked with the reiterative function of the basal ganglia. This fits smoothly with his theory of language evolution, strongly based on sensory-motor control. Like in his last contributions, Lieberman argues in favor of leaving behind the classical Broca–Wernicke model, since it is inaccurate and does not fit with current neuropsychological data, which show that aphasias always present subcortical damage (not just cortical), often in the basal ganglia.

An additional argument in favor of his approach, argues Lieberman, are the new pieces of information available on the FOXP2 gene (p. 171). Lieberman is clearly interested in the function of *FOXP2*, since it has a strong relationship with the control of orofacial muscles and, so it seems, with a decrease of the affected person's IQ. Until very recently, it was believed that *FOXP2* (i.e. the human version of the gene) was a recent innovation in the modern human genome, but according to Krause *et al*. (2007) it was part of the Neandertal genome too — and if so, it was also part of the ancestor's genome of both *H. sapiens* and *H. neanderthalensis* — while a second independent analysis cast a shadow of doubt, since it obtained different results, and the conclusions were that Krause and his collaborators' analysis were contaminated (Coop *et al*. 2008), a new analysis of the gene with a new methodology show new results that are in favor of the presence of the *derived* version of the gene in both Neandertals and modern humans (Burbano *et al.* 2010). In other words, language evolution theories based on motor control arguments like Lieberman's or Corballis' (see above) should take this important factor into account when inferring any relation between control of speech and language. In this respect, Lieberman can overcome this potential theoretical problem — still unknown at the time of the conference — saying that a modern superior vocal tract and the modern speech producing anatomy is present in *H. sapiens* only (p. 173).

A final argument Lieberman provides is the problem of *choking,* which affects every modern human being. According to the author, this problem must have a trade-off, otherwise — as in any adaptationist theory — "there would have been no reason for retaining the mutations that resulted in a human S[uperior] V[ocal] T[tract], unless the neural mechanisms that confer the reiterative properties of speech were in place" (p. 174). But there are many things in the biological evolution of organisms that will always scape from our (human) "logic" way of reasoning if we always think in terms of trade-offs. For example, the presence of the totally useless appendix, whose inflammation will affect the 7% of the world population, according to Brunicardi *et al*. (2004); or wisdom teeth, absent in a low percentage of fortunate people only, provoke more troubles than trade-offs, and they are still there.

In any case, Lieberman is one of the few theoreticians of language evolution who offers a hypothesis which takes into account not only purely theoretical linguistic arguments, but also data from neuropsychology, paleoanthropology, and evolutionary studies like genetics. One may or may not agree with his adaptationist view of language evolution, but one must admit that Lieberman has built a very strong, well-grounded hypothesis, which fits very well with current empirical data from many scientific fields.

The twelfth chapter is devoted to language acquisition and genetics. **Karin Stromswold** presents the results of the work she has been doing on language acquisition and genetics. In short, she compares the heritability factor (h^2) of language in a population group of twins. Assuming that an organism is the result of the phenotypic expression of its genes in an environment and that this process is partially mediated by both perinatal and post-natal environments, the comparison of homozygote vs. heterozygote twins is useful to determine the influence of the environment. Following this procedure, researchers can focus on the heritable genetic factors only. It is a fact that all typically developing humans acquire the basic morphosyntax of their language, "but perhaps some adults fail to master rare linguistic constructions" — like some examples we find in the technical linguistic literature.

Traditional texts on evolution theory talk about the fitness of some traits as the driving force in organic evolution. When Stromswold and colleagues look at the possible relationship between greater linguistic precociousness or proficiency and reproductive success, they find several interesting results: Consistent with a genetic "stoppage",[10] Stromswold's study shows that firstborns with more siblings are less likely to be language-impaired than latter-born children or children with fewer siblings. In other words, language proficiency could have been seen as something qualitatively important when mating took place. However, Stromswold recalls that previous studies (Alwin 1991) found even more interesting results, and contrary to the reproduction success prediction, "children's vocabulary, verbal SAT and IQ scores are inversely correlated with the number of siblings and spacing of siblings" (p. 179). And finally, it is known that women with more education have fewer children, and later than other women. In my opinion, the lesson here is that a theory of language evolution cannot be built on fitness arguments only. These are for sure important and probably have played a role in it. But it would be all but accurate to exaggerate its role in a process of high complexity as organic evolution.

This is a difficult area of research, since participants are not easy to find. In addition, Stromswold and collaborators had to determine which aspects or components of language should be the targets, since "one cannot merely determine the heritability for overall language" (p. 177). The selected targets were syntax, phonology, and lexicon. Although the author does not even mention it, at first sight, this procedure has a ring of *modularism* à la Fodor (1983), which could surprise scholars who don't feel so comfortable with this hypothesis. Notwithstanding, I think it could also be considered as an indirect test for a strong *modularist* view of language. Stromswold and collaborators found significant genetic overlap for these components, indicating that, *possibly*, "some of the same neural circuitry is necessary" for the smooth running of two or more of those components. Stromswold and colleagues wonder whether those components co-evolved or are partially parasitic on others. However, as Stromswold warns, "it could just be happenstance" (p. 178). Therefore, Stromswold's team has carried out a *Perinatal Environment and Genetic Interaction* study. They employed an enormous amount of linguistic and non-linguistic data, information that covers extensive periods of the twins' lives. The results suggest that there is a high genetic overlap for language and oral motor skills as well as fine motor skills (p. 185). Their interpretation is that this "could reflect shared neural circuitry for tasks that require complex motor control". But overlaps do not end here: linguistic and social abilities also overlap. Again, a plausible explanation about shared neural circu-

10 A conscious family planification due to evident genetic impairments. Thus, families with such impairments would have had less descendents.

itry is offered. Next overlap is even more remarkable and surely will catch the reader's attention: Stromswold and collaborators found a large genetic overlap for phonology and syntax scores, greater than for lexical and either syntax or articulation scores.

Thus far, the whole evolutionary picture has become even more difficult to draw. The enormous task carried out by Stromswold and colleagues is more than welcome to a scientific discipline, *biolinguistics*, which is in need of this kind of research in order to revise the general theory along with empirical data. Otherwise, it would be doomed to an endless dialectical spiral of arguments — something useful by itself, for sure, but only to some degree. Stromswold invites us to rethink HCF's hypothesis about FLN. Two options suggest themselves in order to explain this large overlap for phonology and syntax: Either HCF is wrong or another element should be included into FLN, an element that participates in both phonology and syntax. It is not the first time we find new applicants for the selective group of FLN; for instance, there are well-grounded reasons to consider the inclusion of *Duality of Patterning* into this set (Rosselló 2006), a feature that precisely shows this dual character. I'm pretty sure that this new perspective of language evolution will be appreciated by readers interested in human evolution, whatever their training, for its potential in constraining linguistic proposals.

Ian Tattersall is a reputed paleoanthropologist, who makes a brief but interesting contribution, summing up what is known about the evolution of the sub-tribe *Hominina* until our days. In anthropology, symbolic thinking is one of the most important concepts they use to refer to modern behavior. Chomsky cites this expression as if it was Tattersall's, though it's not. Maybe what it is original from him is making it equivalent to modern language. Anyway, the phrase is sometimes used as certainty of having language, something really controversial (see also my above remarks on Gärdenfors & Osvalth). The hypothesis for human evolution presented by the author is the so-called *Out-of-Africa*, the opposite of the *multiregional* theory. The former tells us that all modern humans are descendents of a small group of African early *H. sapiens*. The latter theory maintains that *H. sapiens* is the result of a continuous interbreeding of different species of the genus *Homo*. More precisely, according to this theory, there was just one species, hence the possibility for interbreeding. Although there are still some scholars defending it, genetics clearly favors the "Out-of-Africa" view. Tattersall also assumes splitting off in two different species, *H. neanderthalensis* and *H. sapiens*. That would mean no interbreeding at all between these two hominins. As I have already noted, it has been found that this was indeed possible (Green *et al*. 2010) but just occasionally (the two species never merged into one), a fact also known as *introgressive hybridization* or simply *introgression*. As already suggested, this fact is relevant for accounts of language evolution that focus on a particular capacity of *H. sapiens*, that do not also pay attention to the presence of that capacity in *H. neanderthalensis*.

In order to expose the emergence of modern cognition, Tattersall embraces the possibility of *co-option* or *exaptation*. Several elements already extant in the mind/brain of our ancestors would have been co-opted, reused for new or additional tasks. Such a mechanism makes 'the work' easier for the emergence of new evolutionary functions. A second mechanism the author takes into account is the so-called *byproduct*, also known as *spandrel* (a term coined by Gould & Lewontin 1979), which is a biological novelty that depends on structural constraints, not on a functional role. The confusion of both mechanism is not uncommon, as if they were the same mechanism.

Thus, Tattersall applies these evolutionary mechanisms to language. In his scenario, speech emerged well before symbolic thinking (as indexed by archaeological record). Language proper and modern cognition would have evolved in *H. sapiens* only. Tattersall reconstructs the evolutionary path of the most representative members of the genus *Homo* through the type of industry each of them is associated with. The association of a determined industry to a species is something difficult and risky because most of the time tools are found in the absence of fossil remains, so that the attribution to a particular species can be a tricky matter. Besides, the author talks about the refinement of the tools and the growth of the brains,[11] arguments which are no more decisive: Recent research on such matters has shown that humans are not special at all; hence, regarding the nerve cell average of their brains, modern humans are equal to other apes and monkeys within the mammalian *order* of primates (Azevedo *et al*. 2009). But primates do stand out when compared with other non-primate species. Thus, relating large brains with intelligence is risky. It seems that the type of interconnection of the different parts of the brain plays a more relevant role — indeed this feature has been put forward to speculate about the possibility that *H. floresiensis* had a kind of modern or at least sophisticated cognition (Falk *et al*. 2005; see also above on Gärderfors & Osvath's chapter). In general, I really think that contributions like Tattersall's should be taken into account in biolinguistics, since they talk about the evolution of the organisms within which, the ancient communication system/ language was embodied. The problem is that the picture of the evolution of primate species continuously changes, as new fossils, archaeological items, and genetic data are gathered, forcing researchers to keep an eye on developments.

Next author is **Derek Bickerton**, who does not beat around the bush: His theory and Chomsky's are incompatible (in spite of Corballis' allegations to the contrary, see above). As we have seen when discussing Chomsky's contribution and HCF, Bickerton reproaches them for saying very little about the paleoanthropological context where language emerged. What's more, he argues that the most crucial questions one could ask Chomsky, Fitch, and Hauser (and advocates) are: (i) how, where, when, and why took place the integration of the elements of $[_{\text{FLB}}$ FLB $[_{\text{FLN}}$ FLN]]; and (ii) why other species close to humans, having some of those elements, had never developed any communication system like ours.

Although the *"how, where, and when"* are logical and legitimate questions, to answer *why* a change took place wherever is nearly impossible to answer. For instance, the process of mutation obeys several factors such as migration, genetic drift, and, in the case of humans, cultural factors that could also play a role. Bickerton's question seems to be grounded on the basis that natural selection is

---

11 The average *Australopithecus'* brain was 450 cc; whereas *H. habilis'* (2.2 mya) was 660 cc, earlier *H. erectus'* (i.e. *H. ergaster's*) was 850 cc, and later *H. erectus* achieved 1,100 cc.

the driving force in evolution. It is true, but it is not the only driving force, as Darwin repeatedly said (see the introduction of this review). The second question also has a deterministic flavor. The evolutionary way followed by different species can be imagined in a three- (if not four-) dimensional space, where species, regarding some traits could be very close, but their routes to achieve that close position where completely different. On the contrary, two species can share several traits and reach different new abilities or features, even if they cohabited the same ecological niche.

More than a half of the text is a criticism of Chomsky's hypothesis, and the rest is devoted to present his own hypothesis. As we have seen, Chomsky's hypothesis has a relative high level of abstraction, which sometimes makes it difficult (to some extent, simply not possible) to adapt to current paleoanthropological and neurobiological empirical data. Maybe, this is a common feature of all testable hypotheses in their first stages. Before introducing his criticism, Bickerton sums up quite well both Chomsky's and HCF's stand regarding the role of recursion in FLN. However, in my view the author fails when he says that "[i]n other words, this 'quite different', fully developed human conceptual system formed a necessary prerequisite for the *emergence* of recursion" (p. 201, emphasis added). Bickerton himself mentions, both at the beginning (p. 199) and at the end of his chapter (p. 210), that according to HCF, recursion could have been used in other domains. Therefore, the conceptual system must not be *per se* a *prerequisite for the emergence* of recursion. The nature of this *emergence* is quite obscure, an aspect never clearly explained by HCF. The particular characteristics of this mechanism has led many to think that, if real, the evolutionary biological explanation could be articulated around the concept of *exaptation* (Tattersall 2004) or even a *spandrel* (Barceló-Coblijn, in press). The conceptual system may be an element that intervened in its emergence — one among others (see Arsenijević & Hinzen's original proposal 2010 for recursion as an epiphenomenon of the interaction of linguistic interfaces) — or it may be not. This point notwithstanding, his revision of the concatenation mechanism regulating anaphoras and sentences (pp. 203–204) invites reflection; though, intuitively speaking, we possibly may find an explanation not so far from that which seems to work for sentences like *Mary saw the man walking to the bus station* (with three possible interpretations), that is, "computationally plausible principles of generation and minimal search" (see Chomsky's contribution, pp. 46–48).

It also deserves mentioning that the vision the author has of language and humans still drags along a strong anthropocentrism, bestowing language and humans the power of "effective command and control over all other species" (p. 200). From a biological point of view, this is, at least, an exaggeration, alas, quite common in linguistics across the board. According to this point of view, Bickerton proposes an adaptationist and gradualist scenario. In his opinion, this proposal is "more consistent with, and can be more readily integrated into, biologically and paleontologically based accounts of the overall process of human evolution" (p. 206). The author, as well as Gärdenfors & Osvath in the sixth chapter, singles out one of the traits of language, *displacement* (see Hockett 1958; though Premack 2004: 303, argues that chimpanzees can also make use of displacement), and considers its essential role in animal communication (like in bees and ants).

He proposes that it could be the propelling force, metaphorically speaking, for ancient hominids to make use of *recruitment*, that is to say, the ability of gathering individuals to reach one target. This is really quite interesting in fact because we see how basic communicative properties are shared by evolutionary very distant organisms. The channel is different, but the use of *displacement* is quite the same. The problems arise when the author tries to integrate his proposal "into, biologically and paleontologically based accounts of the overall process of human evolution". According to Bickerton, the process of language evolution would have begun more or less 2 mya. On the one hand, the author does not designate any species: At that time, we can find *H. habilis*, early *H. ergaster* (1.8 mya) or even the *Parathropus*, as mentioned above. Bickerton cites the Oldowan tools, the first made 2.7 mya, which have been associated to *H. habilis* under statigraphic arguments only[12] (Tattersall points out too the uncertain authorship of such tools, p. 195). Other *H. habilis* remains have indeed been found together with such Industry. Up to here, although Bickerton has moved the discussion at least 2 mya back into the past, it is still not clear which hominid could uniquely satisfy his contention (given the more than probable cohabitation of several species of hominids and great apes).

Another argument that the author borrows to make possible the integration of his hypothesis with paleoanthropological studies relates to the evidence of consumption of carcasses of mega-fauna by hominids of that time, as "the richest source of food", required to sustain greater demands of energy. However, according to the evidence, this activity was occasional and fortuitous. To see in the carcasses the "richest source of food" is to overlook the many resources those hominids had, as recollection, micro-fauna hunting and other daily available rich protein sources (e.g., termites and other insects). As in other proposals about language evolution, too much weight has been put onto a single argument or force (in this case, *recruitment*) by Bickerton.

In the final part of his contribution, Bickerton compresses a lot of information and arguments that have recently been put into question, like the well-known argument over brain growth (also used by Chomsky and other authors in this volume) — which should be reconsidered in light of recent findings about the number of neurons in primate brains (see Tattersall's discussion, above).

Finally, the reader will find **Paul Bingham**'s essay. Bingham is a molecular and evolutionary biologist, and he has contributed a new perspective as well as a new theory of language evolution. Like Bickerton, this author strongly trusts, right from the start, in natural selection as the explanatory mechanism for language (p. 211). The starting point is the classic adaptationist stand: "[V]arious constraints impose adaptive trade-offs, resulting in elite execution of one task at the expense of merely serviceable (or negligible) capacity for another" (p. 211). In other words, in order to develop language, modern humans have lost something along the way. As any other adaptationist theory, it focuses on one element that is

12 See footnotes 3 and 4. These tools were found tens of kilometers away of the next hominid, and this was not a member of the genus *Homo*, but the *Australopithecus garhi*. The controversy still remains and has been revived by recent evidence of tool-use for eating with a datation of 3.39 mya (McPherron *et al*. 2010).

raised to the category of propelling force of the process. This time, it is the "conflicts of interest" that gets center-stage, on the grounds that even "exchange of information is apparently directly determined by conflicts of interest" (p. 212). The author continues: "Design information builds organisms. Organisms replicate this design information by replication" (pp. 212–213). This process generates competition and alternative forms are lost along the way. However, we do not know whether or not some of those lost forms could have been slightly better to some extent and their loss due to factors other than direct competition like, for instance, by chance. Moreover, Bingham assumes that organisms use two strategies to assist the mechanism of replication: personal reproduction and assistance of close kind. Additionally, there is the reflection of some basic ideas from Hamilton (1964a, b) as well as Dawkins' (1990) theory of the 'selfish gene'. In Bingham's view, DNA, though unconscious, is intervened by "Natural Selection so that it builds organisms that tend to behave exactly as they would *if* they were controlled by genetic design that *did* have such conscious interests" (p. 213). Hence, it seems we have the basics of the adaptationist recipe.

Bingham's approach is more connected with the study of communication in a broad sense than with the study of the faculty of language *H. sapiens* developed. His view is that non-human animals "arguably parse highly dynamic, hierarchically nested combinatorial information sets of stupendous complexity" (p. 215), as in combinatorial movements. This is enough for the author to think that they really have all the requirements for language, "but in a more modest scale", since "no other factor than the solution of conflicts of interest problem nor new capability […] needs precede evolution of symbolic communication" (p. 215). According to this theory, there are two kinds of actors: *cooperators*, which are non-kin individuals that cooperate until exceeding the costs of cooperation, and *free-riders*, who fail to pay the initial cost of cooperation and hence cooperation does not evolve. Bingham then develops a theory in which these two factors interact in such a way that they have to solve their own conflicts of interest, which inevitably affect all of them, as it happens in any social network.

The key, for Bingham, resides in the mastery of "elite projection of conspecific threat remotely", which would have appeared within our ancestors. This kind of remote threat produces an "enormous reduction in costs" (pp. 218–219). Bingham's particular effort of integration of his theory with anthropological data focuses on the ability of elite throwing objects that only *H. sapiens* seem to have developed. One could immediately argue that, if this were true, there appeared a new ability (contrary to what the author said at the beginning of the paper). But this would not be completely true, since other apes and monkeys can throw objects somehow. The critical point is degree of mastery. Humans have an expertise on these matters, whereas non-human animals roughly throw whatever they can throw. This improvement would have required, by 2.3 mya (i.e. the period of convergence of several *Hominina*; see footnotes 3, 4 and 12), the redefinition of shoulder, pelvis, and the foot — a process completed roughly by 1.8 mya, when *H. erectus* appeared.

In favor of his theory counts the fact that subsequent experimental work has indeed paid attention to Bingham's hypothesis (among others), and it seems that there are reasons to believe that rhesus monkeys do understand the threat that implies the throwing action by humans, and therefore these researchers have concluded that "the capacity to throw did not co-evolve with psychological mechanisms that accompany throwing; rather, this capacity may have built upon pre-existing perceptual processes" (Wood *et al*. 2008: 360).

In any case, the jump from this kind of ability to language seems to me excessive, to say the least, to be accounted by only one factor (*conflict of interest*). The simplest fact in biology seems to obey more than one factor. Reducing so much the explanatory elements gives us no clue about the emergence of psychological and neuropsychological capabilities that really differ among primates, or even among mammals and other *orders*. Even more so when the concept of language is so diffuse and confused with speech, as in Bingham (p. 221). If there is anything that finally has been differentiated in linguistics, it would be the core concepts of *speech* and *language*. What is surprising is that, once we have seen such fuzzy use of these concepts, Bingham accuses linguists of ignoring the very famous *H. erectus* endocast in which Broadfield *et al*. (2001) argued to have finally detected an incipient modern form of the Broca's cap (p. 222). Besides the contentious current status of the classic Broca–Wernicke model (see above, on Lieberman's chapter), I'm sure that scientists like Falk (2007) or Lieberman, who have dealt with this kind of empirical paleo-data and have enormously contributed to the understanding of language evolution, would have a say on this. Although I concede that not every linguist knows about it, maybe the issue is, once again, the concept of a 'linguist' different scientists may have in mind.

As I said at the beginning of this review, there are as many theories on language evolution (in a very brief period of time) as there are people. In this volume alone, there is almost an original hypothesis for each author (certainly, Hauser, Chomsky, and Fitch agree on the basics, but some differences can also be detected). Contrary to appearances, this proliferation of hypotheses could turn out to foster a new, synthetic approach more in tune with current paleoanthropology, anthropological genetics, and evolutionary biology.

I really think that language evolution theory has to leave behind this obsession in finding the element that makes humans special (the key factor, in Lieberman's words). The combination of elements that make us humans human is that key factor. This is neither popular nor spectacular, but it is more in tune with current evolutionary studies. If scholars more or less agree that language is a complex object, then let's think about it in complex ways, taking into account as many variables as possible in order to enrich the general picture. Neither the scientist who proposed natural selection as an evolutionary force nor modern studies on evolution and development ever stated that this is the only mechanism of evolution. The integration of Evo-Devo ideas into the biolinguistic field is warmly welcome, but they have to be integrated within the general evolutionary theory of species (hominids and primates, in particular) — along with a well-grounded linguistic theory, as Bickerton and, above all, Lieberman have tried. Much the same can be said as regards neurobiological theories of language evolution. It is true that there is still much to discover about ourselves and our hominid past, but this should not prevent us from aiming this synthesis — quite the opposite.

References

Alwin, Duane F. 1991. Family of origin and cohort differences in verbal ability. *American Sociological Review* 56, 625–638.

Arsenijević, Boban & Wolfram Hinzen. 2010. Recursion as a human universal and as a primitive. *Biolinguistics* 4, 165–173.

Azevedo, Frederico A. C., Ludmila R. B. Carvalho, Lea T. Grinberg, José Marcelo Farfel, Renata E. L. Ferretti, Renata E. P. Leite, Wilson Jacob Filho, Roberto Lent & Suzana Herculano-Houzel. 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *The Journal of Comparative Neurology* 513, 532–541.

Baker, Mark C. 2002. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.

Barceló-Coblijn, Lluís. In press. Evolutionary scenarios for the emergence of recursion. *Theoria et Historia Scientiarum: International Journal for Interdisciplinary Studies*.

Bickerton, Derek. 2007. Language evolution: A brief guide for linguists. *Lingua* 117, 510–526.

Broadfield, Douglas C., Ralph L. Holloway, Kenneth Mowbray, Adam Silvers, Michael S. Yuan & Samuel Màrquez. 2001. Endocast of Sambungmacan 3 (Sm 3): A new Homo erectus from Indonesia. *The Anatomical Record* 262, 369–379.

Brunicardi, F. Charles, Dana K. Andersen, Timothy R. Billiar, David L. Dunn, John G. Hunter, Jeffrey B. Matthews & Raphael E. Pollock. 2004. *Schwartz's Principles of Surgery*, 8th edn. New York: McGraw-Hill Professional.

Burbano, Hernán A., Emily Hodges, Richard E. Green, Adrian W. Briggs, Johannes Krause, Matthias Meyer, Jeffrey M. Good, Tomislav Maricic, Philip L. F. Johnson, Zhenyu Xuan, Michelle Rooks, Arindam Bhattacharjee, Leonardo Brizuela, Frank W. Albert, Marco de la Rasilla, Javier Fortea, Antonio Rosas, Michael Lachmann, Gregory J. Hannon & Svante Pääbo. 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. *Science* 328, 723–725.

Chittka, Lars & Jeremy Niven. 2009. Are bigger brains better? *Current Biology: CB* 19, R995–R1008.

Chomsky, Noam. 1975. *Reflections on Language*. New York: Pantheon.

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

Chomsky, Noam. 2008. On phases. In Robert Freidin, Carlos P. Otero & Maria L. Zubizarreta (eds.), *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud*, 133–166. Cambridge, MA: MIT Press.

Chomsky, Noam, Marc D. Hauser & W. Tecumseh Fitch. 2005. Appendix: The minimalist program. Ms., Cambridge, MA: MIT and Harvard University. [http://www.wjh.harvard.edu/~mnkylab/publications/recent/EvolAppendix.pdf, (25 October 2010).]

Coop, Graham, Kevin Bullaughey, Francesca Luca & Molly Przeworski. 2008. The timing of selection at the human FOXP2 gene. *Molecular Biology and Evolution* 25, 1257–1259.

Corballis, Michael C. 2002. *From Hand to Mouth: The Origins of Language*. Princeton, NJ: Princeton University Press.

Corballis, Michael C. 2004. The origins of modernity: Was autonomous speech the critical factor? *Psychological Review* 111, 543–552.

Culicover, Peter W. & Ray Jackendoff. 2005. *Simpler Syntax*. New York: Oxford University Press.

Dawkins, Richard. 1990. *The Selfish Gene*, 2nd edn. Oxford: Oxford University Press.

Falk, Dean. 2007. The Evolution of Broca's Area. *IBRO History of Neuroscience* [http://www.ibro.info/Pub/Pub_Main_Display.asp?LC_Docs_ID=3145 (25 October 2010)].

Falk, Dean, Charles Hildebolt, Kirk Smith, Mike J. Morwood, Thomas Sutikna, Peter Brown, Jatmiko, E. Wayhu Saptomo, Barry Brunsden & Fred Prior. 2005. The brain of LB1, Homo floresiensis. *Science* 308, 242–245.

Fitch, W. Tecumseh. 2000. The evolution of speech: A comparative review. *Trends in Cognitive Sciences* 4, 258–267.

Fitch, W. Tecumseh & Marc D. Hauser. 2004. Computational constraints on syntactic processing in a nonhuman primate. *Science* 303, 377–380.

Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179–210; discussion 211–225.

Fodor, Jerry. A. 1983. *The Modularity of Mind: An Essay on Faculty Psychology*. Cambridge, MA: MIT Press.

Gentner, Timothy Q., Kimberly M. Fenn, Daniel Margoliash & Howard C. Nusbaum. 2006. Recursive syntactic pattern learning by songbirds. *Nature* 440, 1204–1207.

Gould, Stephen Jay & Richard Charles Lewontin. 1979. The spandrels of San Marco and the Panglossian paradigm: A critique of the adaptationist programme. *Proceedings of the Royal Society of London. Series B, Containing Papers of a Biological Character* 205, 581–598.

Green, Richard E., Johannes Krause, Adrian W. Briggs, Tomislav Maricic, Udo Stenzel, Martin Kircher, Nick Patterson, Heng Li, Weiwei Zhai, Markus Hsi-Yang Fritz, Nancy F. Hansen, Eric Y. Durand,Anna-Sapfo Malaspinas, Jeffrey D. Jensen, Tomas Marques-Bonet, Can Alkan, Kay Prüfer, Matthias Meyer, Hernán A. Burbano, Jeffrey M. Good, Rigo Schultz, Ayinuer Aximu-Petri, Anne Butthof, Barbara Höber, Barbara Höffner,Madlen Siegemund, Antje Weihmann, Chad Nusbaum, Eric S. Lander, Carsten Russ, Nathaniel Novod, Jason Affourtit, Michael Egholm, Christine Verna, Pavao Rudan, Dejana Brajkovic, Zeljko Kucan, Ivan Gušic, Vladimir B. Doronichev, Liubov V. Golovanova, Carles Lalueza-Fox, Marco de la Rasilla, Javier Fortea, Antonio Rosas, Ralf W. Schmitz, Philip L. F. Johnson, Evan E. Eichler, Daniel Falush, Ewan Birney, James C. Mullikin, Montgomery Slatkin, Rasmus Nielsen, Janet Kelso, Michael Lachmann, David Reich & Svante Pääbo. 2010. A draft sequence of the Neandertal genome. *Science* 328, 710–722.

Hamilton, William D. 1964a. The genetical evolution of social behaviour. I. *Journal of Theoretical Biology* 7, 1–16.

Hamilton, William D. 1964b. The genetical evolution of social behaviour. II. *Journal of Theoretical Biology* 7, 17–52.

Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.

van Heijningen, Caroline A.A., Jos de Visser, Willem Zuidema & Carel ten Cate. 2009. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences of the United States of America* 106, 20538–20543.

Hockett, Charles F. 1958. *Course in Modern Linguistics* (later printing). New York: Prentice Hall.

Jackendoff, Ray. 2002. *Foundations of Language: Brain, Meaning, Grammar, Evolution*. New York: Oxford University Press.

Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser, and Chomsky). *Cognition* 97, 211–225.

Johansson, Sverker. 2005. *Origins of Language: Constraints and Hypotheses (Converging Evidence in Language and Communication Research)*. Amsterdam: John Benjamins.

Kochanski, Greg. 2004. Comment on Fitch & Hauser (2004). Ms., Oxford: Oxford University. [http://kochanski.org/gpk/papers/2004/FitchHauser (25 October 2010).]

Krause, Johannes, Carles Lalueza-Fox, Ludovic Orlando, Wolfgang Enard, Richard E. Green, Hernán A. Burbano, Jean-Jacques Hublin, Catherine Hänni, Javier Fortea, Marco de la Rasilla, Jaume Bertranpetit, Antonio Rosas & Svante Pääbo. 2007. The derived FOXP2 variant of modern humans was shared with Neandertals. *Current Biology: CB* 17, 1908–1912.

Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler & Michael Studdert-Kennedy. 1967. Perception of the speech code. *Psychological Review* 74, 431–461.

Lieberman, Philip. 1999. Silver-tongued Neandertals? *Science* 283, 175.

Lieberman, Philip. 2002. On the nature and evolution of the neural bases of human language. *American Journal of Physical Anthropology* 119, 36–62.

Lieberman, Philip & Edmund S. Crelin. 1971. On the speech of Neanderthal man. *Linguistic Inquiry* 2, 203–222.

Lovejoy, C. Owen. 2009. Reexamining human origins in light of Ardipithecus ramidus. *Science* 326, 74e1–8.

Lovejoy, C. Owen, Bruce Latimer, Gen Suwa, Berhane Asfaw & Tim D. White. 2009. Combining prehension and propulsion: The foot of *Ardipithecus ramidus*. *Science* 326, 72e1–8.

Lovejoy, C. Owen, Gen Suwa, Linda Spurlock, Berhane Asfaw & Tim D. White. 2009. The pelvis and femur of *Ardipithecus ramidus*: The emergence of upright walking. *Science* 326, 71e1–6.

McPherron, Shannon P., Zeresenay Alemseged, Curtis W. Marean, Jonathan G. Wynn, Denné Reed, Denis Geraads, René Bobe & Hamdallah A. Béarat. 2010. Evidence for stone-tool-assisted consumption of animal tissues before 3.39 million years ago at Dikika, Ethiopia. *Nature* 466, 857–860.

Ott, Dennis. 2007. Reverse-engineering the language faculty: Origins and implications of the Minimalist Program. *Harvard Working Papers in Linguistics* 12, 77–90.
Perruchet, Pierre & Arnaud Rey. 2005. Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin & Review* 12, 307–313.
Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95, 201–236.
Plummer, Thomas. 2004. Flaked stones and old bones: biological and cultural evolution at the dawn of technology. *American Journal of Physical Anthropology, Supplement* 39, 118–164.
Premack, David. 2004. Psychology: Is language the key to human intelligence? *Science* 303, 318–320.
Rauschecker, Josef P. & Sophie K. Scott. 2009. Maps and streams in the auditory cortex: Nonhuman primates illuminate human speech processing. *Nature Neuroscience* 12, 718–724.
Rosselló, Joana. 2006. Combinatorial properties at the roots of language: Duality of patterning and recursion. In Joana Rosselló & Jesús Martín (eds.), *The Biolinguistic Turn: Issues on Language and Biology*, 162–186. Barcelona: PPU.
Sperber, Dan & Deirdre Wilson. 1995. *Relevance: Communication and Cognition*. Oxford: Blackwell.
Seyfarth, Robert M. & Dorothy L. Cheney. 2003. Signalers and receivers in animal communication. *Annual Review of Psychology* 54, 145–173.
Susman, Randall L. 1988. Hand of Paranthropus robustus from Member 1, Swartkrans: Fossil evidence for tool behavior. *Science* 240, 781–784.
Tattersall, Ian. 2004. What happened in the origin of human consciousness? *Anatomical Record, Part B: New Anatomist* 276, 19–26.
Tomasello, Michael. 1999. *The Cultural Origins of Human Cognition*. Cambridge, MA: Harvard University Press.
Wood, Justin N., David D. Glynn & Marc D. Hauser. 2008. Rhesus monkeys' understanding of actions and goals. *Social Neuroscience* 3, 60–68.

Lluís Barceló-Coblijn
Group of Human Evolution and Cognition (EVOCOG)
Universitat de les Illes Balears
Departament de Filosofia i Treball Social
Edifici Guillem Cifre de Colonya
carretera de Valldemossa km. 7,5
E-07122, Palma (Mallorca)
Spain
lluis.barcelo@uib.cat

⋆ FORUM ⋆

Lenneberg's Views on Language Development and Evolution and Their Relevance for Modern Biolinguistics

Cedric Boeckx & Víctor M. Longa

1. Introduction

Among the early pioneers of the biolinguistic enterprise (on which see Jenkins 2000, 2004, and Di Sciullo & Boeckx 2011), the names of Noam Chomsky and Eric Lenneberg stand out. Both did more than anyone else to make the study of language a biological topic. They did so in different, complementary ways (ways that we think are beginning to converge in a productive fashion for a variety of factors which we will not expand on here; cf. Boeckx 2010, Di Sciullo *et al*. 2010, Boeckx *et al*. 2011, Balari *et al*. 2011, and Di Sciullo & Boeckx 2011 for discussion). Chomsky stressed the importance of certain basic facts such as the creative aspect of language use and the poverty of the stimulus the child receives during language acquisition to call for the study of the innate factors underlying language growth and to bridge the gap between the tacit knowledge of language users and the primary linguistic data. Lenneberg provided arguments that were much closer to 'wet' biology.[1] In so doing, Lenneberg provided the first and to this day one of the clearest examples of what Boeckx & Grohmann (2007) dubbed "biolinguistics in the strong sense", a body of work of the highest interdisciplinary quality.

In fact, Lenneberg (1967: vii) started from the claim that, as regards language, "biology has been badly neglected". Like Chomsky, his intention was to "reinstate the concept of the biological basis of language capacities" (p. viii). He developed that aim by approaching language as a species-specific mental organ with non-trivial biological properties, which grows in the mind/brain of the

The preparation of the present work was supported by a Marie Curie International Reintegration Grant from the European Union (PIRG-GA-2009-256413; Recipient: Cedric Boeckx), research funds from the Universitat Autònoma de Barcelona Vice-Rector for Research, as well as grants from the Spanish Ministry of Science and Innovation (FFI-2010-20634; PI: Cedric Boeckx) and from the Generalitat de Catalunya (Grant 2009SGR1079 to the Centre de Lingüística Teòrica). Víctor M. Longa was funded by the Spanish Ministry of Science and Innovation and FEDER under the Project "Biolinguistics: Evolution, development, and fossils of language" (FFI2010-14955). We are grateful to two anonymous *Biolinguistics* reviewers for their comments and suggestions on this paper.

[1] Lenneberg too was concerned with promoting nativism, as the following passage reveals: "There is, then, nothing unscientific about the claim that a species-specific behavior pattern, such as language, may well be determined by innate mechanisms" (Lenneberg 1967: 221).

ISSN 1450–3417
Biolinguistics 5.3: 254–273, 2011
http://www.biolinguistics.eu

child in the same way that (other) biological organs grow, showing that the child's path to language displays the hallmark of biological growth.

Lenneberg's (1967) book *Biological Foundations of Language* is today regarded as a classic. Like all classics, it deserves to be re-read at regular intervals, not only to appreciate the success (and limitations) of previous attempts at a synthesis among fields, but also to learn things that we all too often forget. It is from this perspective that we decided to go back to Lenneberg's seminal work. Not to stress its importance, for this is already well established in the literature, but rather to make the point that Lenneberg's conception of the biology of language was much more modern than some more current conceptions, and in fact much more modern than one ought to have expected from a work written in the 1960s, in the heyday of Modern Synthesis in biology.

For such an objective to be fulfilled, we have chosen two topics which as far as we know, have remained unrecognized, or, at least, have not received the attention they deserve in the context of the assessment of Lenneberg's legacy. The first one is Lenneberg's treatment of development and related issues, especially the role he attributed to genes; the second is his treatment of the issue of domain specificity (or lack thereof) in the context of language. Our choice is not altogether innocent: These two areas are at the forefront of current biolinguistics, having been highlighted in the context of the *FOXP2* discovery and of the Hauser *et al*. (2002) Faculty of Language Narrow / Broad distinction.

Our aim is to show that Lenneberg's book has more merits than those usually attributed to it. He did not merely call for an explicitly biological approach to the study of human language at a crucial time in the development of cognitive science; he did so with really modern, indeed prescient, ideas and with 'biological' intuitions that the new biology is beginning to make standard.

2. Lenneberg on Genes, Development, and Maturation

Genes are undoubtedly an important piece of the organismal biological machinery, and they unquestionably play a role in developmental processes. These claims are near truisms. But how relevant are genes for such developmental processes to be fulfilled? Currently, a strong disagreement exists on that issue, and opposite answers are being offered. The "developmentalist challenge" (Weber & Depew 2001), which we are about to expand on, has criticized and undermined the Neo-Darwinian geno-centric stance that has defined modern biology for the past half-century.

Several decades ago, things were different. When Lenneberg wrote his book, biology was almost entirely dominated by the Neo-Darwinian postulates, which can be briefly summarized as the claim that genes are the only relevant materials for explaining development and evolution. Lenneberg's thoughts went beyond this 'orthodox' conception. Unfortunately, his views on this matter were never, as far as we know, acknowledged nor highlighted. Worse, we will see below that they have even been mischaracterized. By means of quotes from his 1967 book, we will try to show that Lenneberg considered genes to be a mere starting-point, which is to be complemented with and related to many biological elements and levels for making up non-trivial developmental paths.

2.1. *Two (Ancient and Modern) Answers about the Generation and Development of Biological Form*

To begin with, let us make clear certain very general positions about development. These will help us situate Lenneberg's vision better. How is biological form generated? Which are the sources for it? How does it develop? These questions, and their many ramifications, have been at the heart of biology for centuries. Accordingly, much ink has been spilt on them, and very disparate answers have been offered. Perhaps one of the clearer examples of the controversy surrounding those questions can be found in the 17th and 18th centuries. At that time, a heated debate arose about how generation and development of form should be considered (for a brief overview, see Maienschein 2005; for an in-depth analysis, see Pinto-Correia 1997 and especially Roe 1981). On the one hand, preformationists believed that the fetus preexisted in the form of an homunculus; although the homunculus was allegedly a being in miniature, it was conceived of as fully formed (i.e. preformed), with organs, limbs, traits, and so on. On the other hand, proponents of epigenesis assumed that the fetus did not preexist at all; instead, it developed step by step.

As biology advanced, it became clear that the preformationist position lacked any kind of empirical support, for the alleged homunculus was never found. However, there is a sense in which the Evolutionary Synthesis and the Neo-Darwinian movement which grew out from it, resurrected the preformationist position.[2] Such a resurrection was carried out by means of a much more subtle strategy, based on the notion of information, and it gave rise to the geno-centric stance, which endows genes with a "special directive, formative, or informative power" (Oyama 2001: 178). Accordingly, genes grew in importance to the point that they were taken to possess the essential information for the development of organisms (and traits) and claimed to be the only relevant elements for heredity, development, and evolution to take place. Hence the return of preformationism, which for example Mayr (1982: 106) explicitly adheres to: According to Mayr, the development of organisms "is controlled by something preformed, now recognized as the genetic program". For those reasons, a strict identification between form and genetic codification was a key premise of Neo-Darwinian thought. In the words of Maynard-Smith & Szathmáry (1999: 2): "[…] each egg contains, in its genes, a set of instructions for making the appropriate adult. […] [I]t is the information contained in the genes that specifies the adult form".

According to such a view, the sources of development lie in the information contained within the genes (nuclear DNA); thus, the conclusion is drawn that development merely consists of displaying something already contained within the genes, by means of a process also strictly directed by them. Hence the preformationism; in fact, Lewontin (2000a: xii) asks "what important difference is there except in mechanical details between a preformed individual and all the

[2] For justification of the preformationist nature of Neo-Darwinism, see Lewontin (2000a), Oyama (2000), Bateson (2001), or Longa (2008), among others. Griffiths & Stotz (2000: 34) coin the term of "neo-preformationism" ("This strategy is often described as 'neo-preformationism' because like the old preformation theory of the embryo it denies that the order manifested in the developed organism actually originates during development"), whereas Weber & Depew (2001: 241) prefer to name it "weak preformationism".

information necessary to specify that individual?". The result would in both cases essentially be the same: "[A]dult organisms are merely expanded versions of the fertilized egg" (Bateson 2001: 156).[3]

Quite the opposite characterizes the developmentalist position (cf. Johnston 1987, Oyama 2000, Gottlieb 2001, Moore 2001, Oyama *et al*. 2001, Johnston & Edwards 2002, Robert 2004, and Blumberg 2005, among many other works). This perspective considers that traits (whether physiological or cognitive) cannot be pre-specified in advance nor directly encoded in the genome. The notion of genetic program as the main (or unique) source of information for development is said to be misguided because it ignores the informational role of many non-DNA elements which are placed between genes and environment, and without which development would not take place. As already pointed out by Lehrman (1953), to assume that the genes have the information for traits, and consequently that traits are encoded or pre-specified in a genetic program, entails turning development into an irrelevant notion.[4]

In fact, development cannot be conceived of as a single function of the genotype, nor as an additive result of genetic and environmental influences alone. The reason is that development is a highly complex process defined by multiple interactions which arise from multiple biological states and stages. Hence the notion of epigenesis, the meaning of which is that "development emerges via cascades of interactions across multiple levels of causation" (Spencer *et al*. 2009: 79), only some of them being related to the genetic level. Accordingly, the traditional divide between genes and environment cannot capture the subtleties of developmental processes: In order to explain traits, much more is needed than genes and environment alone (cf. Johnston 1987, and section 2.3 below).

To sum up, the aforementioned positions are very different perspectives about form and development: According to preformationism, genes do have traits (or, equivalently, the information for them to be generated), whereas from an epigenesist perspective genes cannot contain traits at all; genes are simply a step (perhaps not even the first; cf. Newman 2010) towards a very complex sequence of events which, at the very end, produces the development of traits.[5]

---

3 It should be noted that the geno-centric perspective also pervades some versions of Evo-Devo (cf. Benítez-Burraco & Longa 2010, Linde Medina 2010, Pigliucci & Müller 2010). For example, according to Carroll (2005: 35), "[i]n the entire complement of DNA of a species (the genome) there exists the information for building that animal. The instructions for making five fingers, or two eyespots, or six legs, or black and white stripes are somehow encoded in the genomes of the species that bear those traits".

4 This is not to say that developmentalism defends "a return to empiricism and notions of a 'blank slate'" (Spencer *et al*. 2009: 84); see Maclaurin (2002) for an even clearer position on that topic, and Lorenzo & Longa's (2009) developmentalist proposal within linguistic minimalism. It just means to say that developmental processes are much more complex than traditionally assumed.

5 Both perspectives can be related to the sophisticated conceptual analysis of the 'gene' notion made by Moss (2003), who distinguishes two types of genes which he calls 'Gene-P' and 'Gene-D'. The 'Gene-P' notion corresponds to genes of the preformationist view, those entities being "determinants of organismic traits and phenotypes" (p. xvii), whereas the 'Gene-D', which can be ascribed to the epigenesist/developmentalist perspective, may be conceptualized as simply one of the developmental resources, among many others. Its role is to provide templates for RNA and protein synthesis, but it "has in itself no determinate relationship to organismal phenotypes" (p. xiv).

For the purposes of our paper, it is important to realize that Generative Grammar has traditionally embraced the preformationist perspective by means of its defense of the need for a genetic program for language (see Longa 2008 and Lorenzo & Longa 2009 for details). Modern nativists assumed that there exists a language-specific "genetically determined initial state" (Chomsky 1980: 233) for explaining language growth in the individual. That initial state, named Universal Grammar, contains the innate linguistic principles for language to develop. It is for that reason that Universal Grammar is fully conflated with the notion of linguistic genotype (Chomsky 1980: 65, Lightfoot 1982: 21, Lightfoot 1999: 52, Anderson & Lightfoot 2002: 22, Lightfoot 2006: 45–46), this notion being defined as "that part of our genetic endowment which is relevant for our linguistic development" (Anderson & Lightfoot 2002: 22).[6] Accordingly, many linguistic properties were considered to directly lie in the genes. Examples along these lines abound. Thus, when discussing the property of structure-dependence, Smith (1999: 173) argues that universal properties of language like such a property "have become encoded in the genes of the children". The same strategy has been defended by many generative scholars, who have ascribed many constraints, principles, etc., to the genotype.[7]

2.2. *Lenneberg's Views on Genes*

As just discussed, from its inception Generative Grammar considered the genotype to be the source of the linguistic form, in much the same way that Neo-Darwinism took the genotype to be the source of biological form. For this reason we find it remarkable that Lenneberg showed a very different conception from the view Chomsky and associates adhered to. A careful reading of Lenneberg (1967) shows that, undoubtedly, Lenneberg ascribes a significant role to genes. He considered them to make up a level of biological organization, of course, much like those developmentalist theorists who point out that genes have been vastly overestimated by 'orthodox' biology but who do not deny that those entities have a role (cf. Griffiths & Gray 2005: 420 as an answer to misunderstandings according to which in developmentalist thinking genes are thought to be unimportant or even irrelevant). But, crucially, for Lenneberg, this was not the only level, nor the unique level which contributes to development with information. As we will see in the quotes that follow, Lenneberg went beyond the simplistic view which linked genes and traits. Importantly, he did so in a period when the traditional geno-centric position was even reinforced by the notion of genetic

---

6 In this context, Jenkins' (1979: 106) characterization of Generative Grammar as belonging to the "traditional study of the genetics of organisms" does not come as a surprise.

7 Here are some others:

> "If innate, language must be genetic." (Uriagereka 2007)

> "It seems a miracle that young children easily learn the language of any environment into which they were born. The generative approach to grammar, pioneered by Chomsky, argues that this is only explicable if certain deep, universal features of this competence are innate characteristics of the human brain. Biologically speaking, this hypothesis of an inheritable capability to learn any language means that it must somehow be encoded in the DNA of our chromosomes. Should this hypothesis one day be verified, then linguistics would become a branch of biology." (Jerne 1993: 223)

program that molecular Neo-Darwinism brought to the fore in the beginning of the sixties.

Here is what we get when we turn to *Biological Foundations of Language*. According to Lenneberg, "genetic mechanisms definitely play a role in the development of an individual's behavior" (p. 22). However, Lenneberg's position on genes is characterized by two main aspects: The acknowledgement of a very indirect relationship between genes and traits, and the rejection of the existence of 'special' genes for language, that is, the rejection of the need for a specifically linguistic genotype. Consequently, he offered well taken ideas (with far-reaching ramifications) about what genes can or cannot do, and what their real function is. These aspects can be summarized, as we will show, in Lenneberg's lucid rejection of the following two assumptions: (i) genes contain traits, and (ii) genes directly determine traits.

Chapter 6 of Lenneberg (1967) is particularly enlightening in order to fully appreciate his views on those issues. There, Lenneberg discusses the role of the genes by analyzing "what is known about the specific action of genes" (p. 239). His answer is that "DNA molecules, the biochemical correlates of genes, probably do not more than control the protein synthesis within the cell". That is, according to Lenneberg, no kind of functional principle could be stored in the genes, a claim which prevents genes from specifying any kind of traits in advance. Let's note how similar that statement is to claims made by defenders of the developmentalist stance; for example, according to Bateson (2001: 157), "[g]enes store information coding for the amino acid sequences of proteins; that is all. They do not code for parts of the nervous system and certainly do not code for particular behavior patterns".

Lenneberg went even beyond this, and explicitly denied in several passages of his book the idea that there exist genes for specific traits (including language), or, put similarly, that genes can contain traits. An example illustrates this. Lenneberg points out that the synthesis and biochemical structure of the enzymes are controlled by the molecular structure of the genes, and small changes in that structure "may easily affect the catalyzing efficiency of the enzymes and thereby change the temporal proportions of many far reaching reactions" (p. 241). Those temporal irregularities may affect aspects like the speed of growth, therefore giving rise to altered temporo-spatial patterns. It is for that reason that "genes may be responsible for the inheritance of certain structural characteristics such as the famous Hapsburg lip,[8] or a shortening of the chin, or excessively long legs" (p. 241). In cases like those, the growth is abnormally scheduled: It continues for a longer time than the usual growth (or it may be inhibited before the usual growth).

Lenneberg goes on to argue, very relevantly, that "it is not strictly correct to speak of genes for long ears, for auditory acuity, *or for the capacity for language*" (p. 241) [italics ours]. The reason is that "[g]enes can only affect ontogenesis through varying the cells' repertoire of differentiation, but this, in turn, may have

[8] The 'Hapsburg lip' is a genetic disorder which consists of a thick and overdeveloped lower lip. That disorder is usually associated to the 'Hapsburg jaw' (mandibular prognathism), in which the lower jaw is projected forward.

secondary effects upon structure, function, and capacities" (p. 241).[9]

True, an anomalous sequence of DNA can cause an abnormal phenotype, but the crucial point is that the mirror image cannot be inferred: The correct version of that DNA sequence cannot be taken to be responsible for the trait. To put it in other words, the existence of a correlation between a given sequence of DNA and the presence of a trait does not entail at all a direct causation of the trait by that sequence of DNA. As Bateson (2001: 157) puts it, "[a] disconnected wire can cause a car to break down, but this does not mean that the wire by itself is responsible for making the car move".

As mentioned above, the assumption of a direct link between genes and traits has pervaded and still does not only society, but continues to populate academic writings, especially in the context of language. For example, the recent finding of the human version of the *FOXP2* gene in Neanderthals (Krause *et al*. 2007) has raised the widely extended inference that *Homo neanderthalensis* had a human-like (i.e. modern) language. This case illustrates pretty well that the assumption of a direct link between genes and traits is still accepted by many scientists, in the light of quotes like the following:

> […] Neanderthals must have had a communication system at least equivalent to the one we can infer for Aurignician moderns. […] This is consistent with recent genetic evidence (Krause *et al*. 2007) indicating that a critical gene known to underlie speech — namely *FOXP2* — was present in the Neanderthal genome. (d'Errico & Vanhaeren 2009: 38)

> Up to date behavioural and anatomical studies of neandertal fossils and the recent discovery of their possession of the *FOXP2* gene indicate Neandertals (and, very likely, their European ancestors) has linguistic capacities similar to living humans. (Frayer *et al*. 2010: 113)

However, the inference '*FOXP2* in Neanderthals, *ergo* (complex, sapiens-like) language' is very simplistic, and therefore, ill founded (cf. Benítez-Burraco & Longa, in press for discussion). Among many other reasons, a gene by itself is useless. As Dick Lewontin likes to stress, DNA is a dead molecule, among the most non-reactive, chemically inert of molecules in the living world. That is why it can be recovered from ancient plants and long-dead animals. It has no power to reproduce itself and, while it is promoted as producing proteins, in fact proteins (enzymes) produce DNA. As Fisher & Scharff (2009: 173) put it, "[i]t is worth emphasizing that because language is clearly underpinned by multifactorial influences, the status of a single gene in ancient DNA is insufficient to resolve long-standing debates over linguistic capacities of our extinct ancestors".

The widely extended assumption of a direct link between genes and traits derives from classical genetics (Mendelian–Morganian-like), which was conceived of as the analysis of discrete units acting upon specific phenotypic traits, and it basically entails considering a gene as a simple causal agent (Jablonka & Lamb 2005: 6). The (false) rationale underlying that reasoning has been nicely unraveled (and criticized) by Brian Goodwin:

---

9 In so doing, Lenneberg solves an apparent paradox that he himself raised (cf. p. 272): If the inherited genetic information has only to do with intracellular events, but at the same time language is a supracellular phenomenon, how could language have a genetic foundation?

> So a change in one gene can make a big difference to the shape of an organism, or indeed to any other inherited property. This is a very important observation, and a lot has been made of it. But the conclusion is often drawn that the genes themselves, through their products, contain the key to understanding how all the detailed properties and structures of organisms are made, so that all we need to know is what the genes are doing in order to explain how organisms get their shapes. […] The logic that leads to this very strong statement runs basically as follows. Because we know that a change in a single gene is enough to cause a change in the structure of an organism, genes must contain all the information for making that structure. If we can get that information, we'll understand how the structure is made. (Goodwin 1994: 16–17)

This conception of the genes as simple causal agents has led to the idea that a given gene is the direct and unique responsible for a given phenotype. That view, we insist, was not assumed by Lenneberg. In fact, he rejected that idea at least in two more passages of his book. Thus, Lenneberg points out that the way how genes influence the general patterns of structure and function is by means of their action upon ontogenesis. And, as he states, "it is possible to talk about language in connection with genetics without having to make shaky assumptions about 'genes for language'" (p. 244). Later, he goes on saying that although pedigrees and twin studies suggest that genetic transmission is relevant to language facilitation, "there is no need to assume 'genes for language'" (p. 265). To have firmly placed that conclusion on the agenda almost half a century ago is undeniably a great merit of Lenneberg's thought, one that too many (bio)linguists continue to ignore.

According to Lenneberg, "we do not know what the direct relationships are between man's complement of genes and his mode of communication; we merely wish to outline the theoretical possibilities for relating the two" (p. 244). As it can be appreciated, Lenneberg outlined those theoretical options (which are based on the existence of a very indirect relationship between genes and traits) very accurately. Lenneberg was perfectly aware that "[t]he idea that there is a gene for adventurousness, heart disease, obesity, religiosity, homosexuality, shyness, stupidity,[10] or any other aspect of mind or body has no place on the platform of genetic discourse" (Jablonka & Lamb 2005: 6).[11]

All the aspects raised so far naturally lead to discuss Lenneberg's view on development in general, and on language development in particular. Again, his treatment of these topics is both illuminating and remarkably modern-sounding.

2.3. Lenneberg on Development

In section 2.1, it was argued that the conception of development held by (ancient

[10] Or even voting behavior. Fowler & Dawes (2008: 587–588) state that "two extensively studied genes are significantly associated with voter turnout. Further, these are the first two genes ever directly associated with political behavior". It seems to us that this view would also be strongly rejected by Jablonka & Lamb (and by Lenneberg).

[11] Critics may argue against our interpretation of some passages of Lenneberg's book using quotes like "[a] direct and profound dependence of language capacity on genetic constitution" (p. 253). However, our discussion shows that "direct and profound" should not be understood as a direct causation at all.

or modern) preformationism can be said to imply that development is undervalued or even neglected, because the information for the traits is considered to pre-exist in the genes: The more relevant we consider the genetic program, the more trivialized development becomes. A key idea of such a preformationist perspective is to assume that genotype and environment are the only two relevant actors for development to take place. However, from a truly developmental view, it is obvious that this reductionist strategy is misguided, for it ignores the vast biological machinery which exists between genotype and environment. Without that machinery, both genotype and environment would be completely useless. This point is especially clear when we consider what is referred to as 'cascading events' in developmental biology (cf. Moore 2001: chap. 4): The development of any trait implies a very complex sequence of events, where "event A causes event B, which causes event C, and so on" (Moore 2001: 69). For sake of exposition, that phenomenon could be represented in a very simplified way with something like this:

$$A \Rightarrow B \Rightarrow C \Rightarrow D \Rightarrow E \Rightarrow F \Rightarrow G \ldots \Rightarrow Z$$

As Moore points out, cascading events resemble a typical 'domino effect', where a domino pushes over another domino, and so on. The relevant point is the following: "[I]t does not seem reasonable to call event A the cause of event Z, because many other events are involved in producing event Z as well" (p. 70). Therefore, cascading events entail that nor genotypes nor environment may have direct effects. However, those who argue for the preexistence of traits within the genes forget that aspect (cf. Johnston & Edwards 2002 for a proposal of a highly articulated developmental model, with many interactants).

Interestingly, Lenneberg did not fall into that trap. A feature pervading his book is the recognition of the enormously complex nature of development, in such a way that after the genes, "a very complex chain of event ensues, until a relatively steady state, called maturity, is reached" (Lenneberg 1967: 240). This implies that, according to Lenneberg, "organisms are not programmed for their behavior by an ex machina force, but they develop a program ontogenetically together with nervous and non-nervous tissues" (p. 4). Two more quotes make this point abundantly clear:

> The central nervous system and other tissues in the body develop simultaneously and influence one another continuously during morphogenesis.
> (Lenneberg 1967: 28)

> Animals develop as an integrated whole including structure, function and behavioral capacities. (Lenneberg 1967: 240)

Lenneberg clearly held an interactionist, dynamic view of development. That view involves the rejection of an encapsulated and self-sufficient conception of the genome.

Actually, Lenneberg's view can be related to a key feature of epigenesis, which is referred to as (complex) causal co-interactionism (cf. Lewontin 2000b, Oyama *et al*. 2001: 2, Robert *et al*. 2001: 955, Robert 2003: 96). In Robert's words:

> Constructive causal interactions in development involve inducing, facilitating, maintaining, and participating in time-sensitive feedback loops at multiple levels within and beyond the developing organism — only some of which might be characterized as gene activation. The interactions comprising organismal development are complex, and their effects are not simply additive. (Robert 2003: 96)

Complex co-interactionism thus means that development cannot be perceived as a linear series of stages, but as a continuous transformation across the overall process, where "biological products are built up, deformed, broken down, distributed or deformed" (Oyama 2000: 133), those changes being the very essence of development.

According to Lenneberg, there exists an immanent schedule of evolvement evolutionary program "in which apparently one set of events sets the stage for a subsequent set, and so on" (p. 313). This entails that both form and function do not pre-exist; rather, they "gradually develop through a process of differentiation" (p. 373), where "[t]he basic plan is based on information contained in the developing tissues" (p. 373); consequently, such a plan is not to be found in the genes themselves.

A brief discussion in the final passages of the book clearly shows the complete rejection of preformationism (cf. p. 380). There Lenneberg argues that sometimes the claim is made that to defend the species-specificity of behavior or to postulate innate factors determining that behavior implies to return to the preformationism of the 18th century. However, he rejects that idea by stating that "[n]othing could be farther from the truth" (p. 380); then he goes on to argue that "the epigenetic doctrine teaches that the adult form is the result of gradual formation of structure through a continuing process of reconstitution of molecules" (p. 380). Lenneberg's conclusion is as follows: "Clearly, our proposal of how language develops in the individual is in no way counter to an epigenetic view" (p. 380).

The aspects under discussion, and the discussion raised in section 2.2, point to Lenneberg's clear preference for a developmental model lacking any hints of preformationism, that is, an opposite model to the one which traditionally characterized Generative Grammar, based on linguistic principles with a content directly ascribed to the genes.

Lenneberg's view on nativism seems to us to fall within what Stich (1975) called the 'dispositional model' of nativism (as opposed to the 'input–output model'). This model defines an innate trait as a property which is determined to appear in a reliable way at a certain point of the developmental process of any member of the species (cf. Maclaurin 2002 as an updating treatment of that model). The dispositional model seems to fit in well with Lenneberg's conception of development. In fact, many references are made across the book to the "regularity in the sequence of appearance of given milestones" (p. 126) as a hallmark for "maturationally controlled emergence of behavior" (cf. pp. 126, 127, 133, 136, 142, 244, 326, 372, etc.). According to that model, innate traits are due to heterogeneous developmental resources, understood as any factor influencing development. Genetic factors are just one of them, but in no sense can they be regarded as the main or unique factors.

The discussion so far seems to us to provide enough evidence that Lenneberg's vision, encapsulated in his 1967 book, is quite different from the more popular representation of his views, as found, for example in the writings of Ken Wexler on "Lenneberg's dream" (Wexler 2003, in press).

Contra Wexler, we think that Lenneberg's dream cannot be characterized by the idea that development or maturation are pre-specified in the genes (cf. above) nor by the assumption that specific constraints are rooted in the genes. Certainly, Lenneberg stressed the importance of the process of internal maturation for language (cf. pp. 126, 139, 142, etc.), and he claimed that "the appearance of language is primarily dependent upon the maturational development of states of readiness within the child" (p. 142). However, for Lenneberg biological growth is not controlled by genes alone, something which is quite the opposite to Wexler's framework. According to Wexler (2003: 13), development "is in central cases taken to be genetically guided" in such a way that "many principles are genetically programmed" (Wexler 2003: 38). For example, his treatment of the UCC, the "Unique Checking Constraint" (Wexler 1998: 59)[12] clearly illustrates that assumption. According to Wexler (1998: 73), the UCC is "part of the genetic program"; therefore, "[t]he genetic system determines that at birth […] the UCC is in place" (Wexler 2003: 40). Wexler (in press: 38) goes on to argue that "[g]iven these results together with the results discussed in this paper, the field is beginning to hone in on which gene or genes control the development of UCC".

It is obvious from these quotes that, according to Wexler, there is a direct link between genes and a linguistic constraint. Wexler (2003: 45) points out that his framework resembles the way how Lenneberg expected language development to behave; hence, "Lenneberg's dream". However, as our discussion showed, the formulation of Lenneberg's dream cannot match the direction Wexler argues for. Lenneberg, we think, would not agree at all with that interpretation of his dream.

3. How Special Are the Language Mechanisms According to Lenneberg?

The second topic we have chosen from Lenneberg (1967) is his answer to the question whether or not language mechanisms are special. This issue especially matters because some parallels can be traced with the issue of the role provided by Lenneberg to genes and development. Let us clarify this point a bit more: Although he endorsed the innate nature of language, Lenneberg departed from the geno-centric perspective that has characterized Generative Grammar for decades. With regard to language specificity or unspecificity, his own thought also departed from the canonical position held by traditional Generative Grammar. In fact, it seems to us to fit in well with more recent approaches to the issue that soften the strongly system-specific stance that has dominated generative (bio)linguistics for so long.

As pointed out in Balari *et al*. (2010) and Di Sciullo *et al*. (2010), for many years, Generative Grammar was centered on the formal singularity and uniqueness of language as opposed to any other instance of cognition or behavior of

[12] The UCC states that "[t]he D feature of DP can only check against one functional category".

non-human animals. This meant that Generative Grammar highlighted the great specificity of the language faculty, which was conceived of as uniquely human (and functioning with uniquely linguistic mechanisms). Accordingly, Chomsky (1968) assumed that language was a capacity lacking true homologues in other organisms, therefore being a problem for Darwinian-based continuity theses.

However, with the beginning of the 21st century, that position has been somehow reformulated, especially since Hauser *et al*. (2002) and their case in favor of a comparative approach to the language faculty. According to that proposal, the faculty of language is not conceived of as a homogeneous whole (in an all-or-nothing style), but as a collection (or 'mosaic') of abilities and capabilities, some of them being shared with non-human animals and other being uniquely human. The divide between faculty of language in the broad sense (FLB) and in the narrow sense (FLN) represents the attempt to shed light on this issue: Many mechanisms related to the systems of thought, and especially, to the sensori-motor system are not uniquely human, but rather they are widely extended among non-human animals. That is not the case, though, of FLN, which is something like a residue of the uniquely human nature of the language faculty, which, by definition, cannot be compared to anything existing in the mind of other species (nor even in other domains of the human mind). A clear sign of the current interest on that topic is the fact that it is currently being much debated (cf. Hauser *et al*. 2002, Pinker & Jackendoff 2005, Fitch *et al*. 2005, and Jackendoff & Pinker 2005; see also Anderson 2004 and Samuels *et al*., in press).

Where does the interest of Lenneberg's approach to the issue lie? The answer is straightforward: In a moment when Generative Grammar, and Chomsky himself, stressed the differences between human language as a whole and animal systems, by stating that no linguistic mechanisms had anything to do with those found in animals, Lenneberg provided us with an incipient comparative method, which led him to the assertion that similarities between humans and non-humans can be found, even for those areas of the language faculty which in current terms would correspond to the FLN. That way, Lenneberg's view is that key mechanisms of language can be related to very ancient animal capacities.

A clarification is in order: What we have just said does not mean that according to Lenneberg language is not species-specific. It clearly is, and this is one of the most recurrent claims in his book. If that were not the case, animals could have access to language, but, as he repeatedly points out, that does not follow. What that claim implies is that the language specificity is based on the modification in humans of ancestral vertebrate mechanisms, a descent-with-modification view that would have pleased Darwin himself (who preferred the notion of descent to that of evolution). An example will help appreciate this. Consider MacNeilage's (1998) treatment of the evolutionary origins of the syllable (we hasten to add that we do not necessarily endorse this view on the syllable, but we think it nicely illustrates what we want to convey). According to MacNeilage (1998), syllabic cyclicity derives from repetitive movements associated to the mammal chewing. This does not involve to defend that the syllable is not a specifically linguistic unit; undoubtedly, it is. However, that unit is built upon ancestral mammal capacities. Something similar applies to Lenneberg's view: Although some of the features that define the essence of language, like phrase-structure or transformations, are specifically linguistic features, they may nevertheless derive from very ancient capacities.

Lenneberg (1967: chap. 6) discusses some differences between animal communication and language, and he argues against the claim that "a straight-line of evolution" (p. 228) may be traced among them. In fact, he rejects two different versions of the continuity thesis: Straight-line evolution of language with only quantitative changes, and straight-line evolution of complexity by stepwise accretion, with missing links (cf. pp. 228–230). Instead, he defends the theory of discontinuity between animal communication and human language, by arguing that such a thesis "is not only biologically acceptable but, in fact, more in line with present theories in developmental biology than the former type theory" (p. 228). However, he clarifies that "[a] discontinuity theory is not the same as a special creation theory. No biological phenomenon is without antecedents" (p. 234). Although to his view those antecedents are not evident, he offers a highly interesting proposal in that regard, which combines shared nature and specificity.

The key of Lenneberg's proposal (p. 336) is that the cognitive mechanisms that underlie the components of language (syntax, semantics or phonology) are based on the processes of categorization, differentiation, and interrelation (differentiation and interrelation being just two aspects of the general process of categorization). What about the categorization process? Is it special to humans, or is it shared? Lenneberg's answer offers no doubts: Categorization is a universal phenomenon in the animal kingdom, although "the categorizations peculiar to language operate through the application of highly species-specific principles" (p. 336).[13] For that reason, Lenneberg states that the cognitive function which underlies language implies the adaptation in humans of an ubiquitous process among vertebrates of categorization and extraction of similarities: "The perception and production of language may be reduced on all levels to categorization processes" (p.374).

It is important to analyze what the process of categorization is for Lenneberg, for authors who would write much later, like Bickerton (1990), reached very similar conclusions. Bickerton states that categorization implies a process of segmentation of reality that in turn points to a process of abstraction of it. In fact, he showed that any category puts together actions, processes, or entities which cannot be taken as equivalent: For example, we can find very different tables, according to their size, raw materials, color, form, etc. However, when we categorize, we abstract away from those differences, and we keep only the constant features, which are mainly abstract (cf. Cohen & Lefebvre 2005 for an in-depth analysis of mechanisms of categorization).

Actually, intensive research with animals has unequivocally shown that they are also able to categorize, that capacity being at the basis of concept formation. As it happens with humans, animal's categorization involves to

[13] Incidentally, note also the modern character of Lenneberg's conception with regard to the recognition of complex processes in the animal mind. It should be kept in mind that in the 1960s the existence of animal minds was in general not recognized, given the predominance of behaviorism. Consequently, the possibility that animals could possess mental processes like categorization was in general not recognized either.

abstractly unify tokens of a concept which can greatly vary from each other, in such a way that an organism "somehow perceives relations of unity between objects that in superficial detail appear quite different" (Bickerton 1990: 92; cf. Harnad 1987). For such a task to be made, complex capacities such as induction, generalization and abstraction are required (cf. Hurford 2007: 27 ff.).[14]

As regards categorization, Lenneberg asserts that "[m]an is no different from other animal" (p. 298). In fact, language mechanisms are to his mind specific applications of mechanisms which are shared. All vertebrates, Lenneberg claims (p. 331), may superimpose categories of functional equivalence to configu-rations of stimuli, in order to give a single type of response to any member of a concrete stimulus category.[15] That way, animals organize the perceptual world through a process of categorization: "[T]here is no formal difference between man's concept-formation and animal's propensity for responding to categories of stimuli" (p. 332), although a substantive difference exist, which is that the overall possibilities of categorization are not the same among species.

As already mentioned above, from the basic process of categorization, two subsequent processes arise which are also shared by non-human animals: Differentiation or discrimination, and establishment of interrelations among categories, that is, to perceive transformations. And, importantly, Lenneberg uses those mechanisms in order to explain basic properties of language:[16] phrase-structure and grammatical transformations.

Phrase-structure arises, according to Lenneberg, by means of the progressive differentiation of categories which are very general in the initial stages of development. Accordingly, he considers that the progressive development of syntagmatic structure is a process of differentiation of grammatical categories (p. 294), in much a similar vein to the differentiation process that is the essence of the semantic component (by the way, it should be noted his impressive defense of a semantic naturalism) or the phonological one. With the differentiation procedure, he explains even main features of language, like recursion or nested dependencies.[17]

As regards transformations, Lenneberg acknowledges that they have played a pivotal role in the characterization of grammar (p. 296). Transformations operate by relating different phrase-markers, and they allow to perceive relationships and affinities between sentences with a very different surface structure, thus leading to the establishment of grammatical, semantic, and phonological connections (pp. 292, 299–300). But, again, far from considering them to be

14 Thus, "[p]ossession of words is not a necessary criterion for identifying possession of concepts" (Hurford 2007: 10).

15 That capacity can be said to be an instance of MacPhail's (1987) level 2 of intelligence. This level goes beyond the connection between a stimulus and a response (level 1), in such a way that different stimuli can be connected, thus presupposing mental representations.

16 This clearly goes beyond language, because "[t]his differentiation process is not confined to language. In fact, it is the hallmark of all development" (Lenneberg 1967: 295).

17 "Both recursiveness and nested dependencies are simply consequences of differentiation or specification". He goes on to argue that "[o]rganization of phrase-structure with the resulting phenomenon of recursiveness and nested dependencies appears as a 'natural phenomenon' once we assume that a ubiquitous process is influencing a specific behavior", although to execute that behavior requires "specific cognitive and thus biological adaptations" (all from Lenneberg 1967: 296).

uniquely human and uniquely linguistic mechanisms, according to Lenneberg, transformations are a ubiquitous process, also derived from categorization itself.

If, as specified above, categorization implies to group configurations of stimuli that are different from each other, transformations imply to recognize the similarity, something which is not unknown for animals either, and which proceeds through the mechanism of interrelation:

> All animals have the ability to group together stimulus configurations which may be physically totally different from each other; however, the animal makes an identical response to certain ones and thus treats them as if they were similar in some respect; we cannot escape the conclusion that for the animal, some similarity exists among such stimuli. (Lenneberg 1967: 298)

Accordingly, the conclusion holds that "all similarities involve transformational processes" (p. 299). That is, where the grouping is made in terms of a categorization, a transformational process exists. To sum up, for Lenneberg, to perceive similarities must be a deeply entrenched process, in such a way that it points to the true nature of behavioral organization (p. 301).

According to Lenneberg, the common aspect of any transformation is an abstract schema (pp. 298–299); for example, the structural similarity between two strings of words transforms audible physical patterns into an abstract schema. That way, transformations operate by translating physical aspects into abstract schemas or representations, and they can be said to be simply the interrelation of categories (p. 335). For that reason, the transformational principle of language seems identical to the cognitive principle underlying the capacity of categorizing behavioral structures in a wide sense.

To sum up, according to Lenneberg, both phrase-structure and transformations are special applications of general models of organization because they "are common to the organization of the behavior of all higher animals" (p. 302). Therefore, according to Lenneberg, the cognitive function underlying language is the adaptation of a ubiquitous process of categorization and extraction of similarities among vertebrates (p. 374).

A final issue is in order: If language mechanisms are instances of mechanisms which pervade animal kingdom, "why is language species-specific?" (p. 302). The answer has to do with the fact that "cognitive processes must be highly adapted biologically" (p. 302). Although categorization is universal among animals, linguistic categorization operates according to principles which are specific of our species. That way, to perceive similarities and relations depends on the different capabilities of organisms for handling transformations (p. 325), those capabilities being biologically constrained.[18] The result is that the nature of categorization must be determined for each species.

Chapter 6 of *Biological Foundations of Language* offers some hints for the ultimate reasons lying behind the specificity of cognitive processes. In that chapter, Lenneberg discusses D'Arcy Thompson's (1917) famous method of transformations based on the superimposition of Cartesian coordinates on differ-

18 In fact, Lenneberg (p. 371) argues that interspecific differences are not only related to differences in peripheral sensorial thresholds; as regards the cognitive organization, a function of upper, main processes, is also involved.

ent animal forms (or animal parts), the distortion of those coordinates generating several existing forms.

Lenneberg goes on to argue that those relationships among mature forms can be accounted for by "changes in growth gradients during ontogeny" (p. 245). Accordingly, those visible transformations derive from 'invisible' molecular transformations, which cause the developmental histories to differ. For that reason, the answer to the issue of language specificity should rest on developmental changes. That is a very nice result: Development as the source of evolution, much in the spirit of Evo-Devo.

4. Conclusions

Lenneberg's *Biological Foundations of Language* can be considered a key reference for the emergence of the Biolinguistic Program. Accordingly, it has exerted a central influence in placing the study of language in a biological context. However, in this paper we have tried to show that the book's merits by far exceed those which are usually given to it. In order to show that, we have looked at Lenneberg's treatment of the role attributed to genes and development, and his view on the issue of domain specificity for language. Lenneberg's answers to both issues are based on surprisingly modern conceptions, which went beyond the usual treatments on language and biology at the time when the book was written, but also strike us as far more modern than standard conceptions in current biolinguistics. Curiously, those conceptions are more in agreement with ideas brought to the fore by the Minimalist Program in linguistics (relativization of the role attributed to genes, and a new look at the issue of language specificity), and by the calls for an extended synthesis in biology (Pigliucci & Müller 2010). In many ways, Lenneberg's book was clearly ahead of its time.

References

Anderson, Stephen R. 2004. *Doctor Doolittle's Delusion: Animals and the Uniqueness of Human Language*. New Haven, CT: Yale University Press.

Anderson, Stephen R. & David Lightfoot. 2002. *The Language Organ: Linguistics as Cognitive Physiology*. Cambridge: Cambridge University Press.

Balari, Sergio, Antonio Benítez-Burraco, Marta Camps, Víctor M. Longa & Guillermo Lorenzo. 2010. La importancia de ser moderno. Problemas de método e ideología en el debate sobre la cognición y la conducta de los Neandertales. *Ludus Vitalis* XVIII, 143–170.

Balari, Sergio, Antonio Benítez-Burraco, Marta Camps, Víctor M. Longa, Guillermo Lorenzo & Juan Uriagereka. 2011. The archaeological record speaks: Bridging Anthropology and Linguistics. *International Journal of Evolutionary Biology*, vol. 2011. doi: 10.4661/2011/382679.

Bateson, Patrick. 2001. Behavioral development and Darwinian evolution. In Oyama *et al*. (eds.), 149–166.

Benítez-Burraco, Antonio & Víctor M. Longa. 2010. Evo-Devo — of course, but which one? Some comments on Chomsky's analogies between the

biolinguistic approach and Evo-Devo. *Biolinguistics* 4, 308–323.
Benítez-Burraco, Antonio & Víctor M. Longa. In press. El papel del ADN fósil en Paleoantropología: *FOXP2*, Neandertales y lenguaje. *Zephyrus. Revista de Prehistoria y Arqueología*.
Bickerton, Derek. 1990. *Language and Species*. Chicago: University of Chicago Press.
Blumberg, Mark S. 2005. *Basic Instinct: The Genesis of Behavior*. New York: Thunder's Mouth Press.
Boeckx, Cedric. 2010. *Language in Cognition: Uncovering Mental Structures and the Rules Behind Them*. Chichester: Wiley-Blackwell.
Boeckx, Cedric & Kleanthes K. Grohmann. 2007. The *Biolinguistics* manifesto. *Biolinguistics* 1, 1–8.
Boeckx, Cedric, María del Carmen Horno Chéliz & José Luis Mendívil Giró (eds.). 2011. *Language from a Biological Point of View: Current Issues in Biolinguistics*. Book ms., ICREA/Universitat Autònoma de Barcelona & Universidad de Zaragoza.
Carroll, Sean B. 2005. *Endless Forms Most Beautiful: The New Science of Evo Devo and the Making of the Animal Kingdom*. New York: W.W. Norton & Co.
Chomsky, Noam. 1968. *Language and Mind*. New York: Harcourt Brace.
Chomsky, Noam. 1980. *Rules and Representations*. New York: Columbia University Press.
Cohen, Henry & Claire Lefebvre (eds.). 2005. *Handbook of Categorization in Cognitive Science*. Amsterdam: Elsevier.
d'Errico, Francesco & Marian Vanhaeren. 2009. Earliest personal ornaments and their significance for the origin of language debate. In Rudolf Botha & Chris Knight (eds.), *The Cradle of Language*, 16–40. New York: Oxford University Press.
Di Sciullo, Anna Maria & Cedric Boeckx (eds.). 2011. *The Biolinguistic Enterprise. New Perspectives on the Evolution and Nature of the Human Language Faculty*. New York: Oxford University Press.
Di Sciullo, Anna Maria, Massimo Piattelli-Palmarini, Kenneth Wexler, Robert C. Berwick, Cedric Boeckx, Lyle Jenkins, Juan Uriagereka, Karin Stromswold, Lisa Lai-Shen Cheng, Heidi Harley, Andrew Wedel, James McGilvray, Elly van Gelderen & Thomas G. Bever. 2010. The biological nature of human language. *Biolinguistics* 4, 4–34.
Fisher, Simon E. & Constance Scharff. 2009. *FOXP2* as a molecular window into speech and language. *Trends in Genetics* 25, 166–177.
Fitch, W. Tecumseh, Marc D. Hauser & Noam Chomsky. 2005. The evolution of the language faculty: Clarifications and implications. *Cognition* 97, 179–210.
Fowler, James H. & Christopher T. Dawes. 2008. Two genes predict voter turnout. *The Journal of Politics* 70, 579–594.
Frayer, David W., Ivana Fiore, Carles Lalueza-Fox, Jakov Radovčić & Luca Bondioli. 2010. Right handed Neandertals: Vindija and beyond. *Journal of Anthropological Sciences* 88, 113–127.
Goodwin, Brian. 1994. *How the Leopard Changed Its Spots: The Evolution of Complexity*. New York: Charles Scribner's Sons.
Gottlieb, Gilbert. 2001. A developmental psychobiological systems view: Early

formulation and current status. In Oyama *et al*. (eds.), 41–54.
Griffiths, Paul E. & Russell D. Gray. 2005. Discussion: Three ways to misunderstand developmental systems theory. *Biology and Philosophy* 20, 417–425.
Griffiths, Paul E. & Karola Stotz. 2000. How the mind grows: A developmental perspective on the biology of cognition. *Synthese* 122, 29–51.
Harnad, Stevan (ed.). 1987. *Categorical Perception: The Groundwork of Cognition*. Cambridge: Cambridge University Press.
Hauser, Marc D., Noam Chomsky & W. Tecumseh Fitch. 2002. The faculty of language: What is it, who has it, and how did it evolve? *Science* 298, 1569–1579.
Hurford, James. 2007. *The Origins of Meaning: Language in the Light of Evolution*. New York: Oxford University Press.
Jablonka, Eva & Marion J. Lamb. 2005. *Evolution in Four Dimensions: Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge, MA: MIT Press.
Jackendoff, Ray & Steven Pinker. 2005. The nature of the language faculty and its implications for evolution of language (Reply to Fitch, Hauser and Chomsky). *Cognition* 97, 211–225.
Jenkins, Lyle. 1979. The genetics of language. *Linguistics and Philosophy* 3, 105–119.
Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press.
Jenkins, Lyle (ed.). 2004. *Variation and Universals in Biolinguistics*. Amsterdam: Elsevier.
Jerne, Niels K. 1993. The Generative Grammar of the immune system. Nobel Lecture, December 8 1984. In Jan Lindsten (ed.), *Nobel Lectures: Physiology or Medicine 1981-1990*, 211–225. Singapore: World Scientific.
Johnston, Timothy D. 1987. The persistence of dichotomies in the study of behavioral development. *Developmental Review* 7, 149–172.
Johnston, Timothy D. & Laura Edwards. 2002. Genes, interactions, and the development of behavior. *Psychological Review* 109, 26–34.
Krause, Johannes, Carles Lalueza-Fox, Ludovic Orlando, Wolfgang Enard, Richard E. Green, Hernán A. Burbano, Jean-Jacques Hublin, Catherine Hänni, Javier Fortea, Marco de la Rasilla, Jaume Bertranpetit, Antonio Rosas & Svante Pääbo. 2007. The derived *FOXP2* variant of modern humans was shared with Neandertals. *Current Biology* 17, 1908–1912.
Lehrman, Daniel S. 1953. A critique of Konrad Lorenz's theory of instinctive behavior. *Quarterly Review of Biology* 28, 337–363.
Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: Wiley.
Lewontin, Richard C. 2000a. Foreword. In Oyama (2000), vii–xv.
Lewontin, Richard C. 2000b. *The Triple Helix: Gene, Organism, and Environment*. Cambridge, MA: Harvard University Press.
Lightfoot, David. 1982. *The Language Lottery: Toward a Biology of Grammars*. Cambridge, MA: MIT Press.
Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution*. Malden, MA: Blackwell.
Lightfoot, David. 2006. *How New Languages Emerge*. Cambridge: Cambridge University Press.

Linde Medina, Marta. 2010. Two "EvoDevos". *Biological Theory* 5, 7–11.
Longa, Víctor M. 2008. Una vision crítica sobre la noción de 'programa genetico' desde la biología y la lingüística: Consecuencias para la conceptualización de la ontogenia del lenguaje. *Verba* 35, 347–385.
Lorenzo, Guillermo & Víctor M. Longa. 2009. Beyond generative geneticism. Rethinking language acquisition from a developmentalist point of view. *Lingua* 119, 1300–1315.
Maclaurin, James 2002. The resurrection of innateness. *The Monist* 85, 105–130.
MacNeilage, Peter. 1998. The frame/content theory of evolution of speech production. *Behavioral and Brain Sciences* 21, 499–511.
MacPhail, Euan. 1987. The comparative psychology of intelligence. *Behavioral and Brain Sciences* 10, 645–656.
Maienschein, Jane. 2005. Epigenesis and preformationism. *Stanford Enclyclopedia of Philosophy*. [http://www.plato.stanford.edu/archives/fall2008/entries/epigenesist/#7] (last accessed 7 June 2011).
Maynard-Smith, John & Eörs Szathmáry. 1999. *The Origins of Life: From the Birth of Life to the Origin of Language*. New York: Oxford University Press.
Mayr, Ernst. 1982. *The Growth of Biological Thought: Diversity, Evolution, and Inheritance*. Cambridge, MA: Harvard University Press.
Moore, David S. 2001. *The Dependent Gene: The Fallacy of 'Nature vs. Nurture'*. New York: Henry Holt.
Moss, Lenny. 2003. *What Genes Can't Do*. Cambridge, MA: MIT Press.
Newman, Stuart A. 2010. Dynamical patterning modules. In Pigliucci & Müller (eds.), 281–306.
Oyama, Susan. 2000. *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd edn. Durham, NC: Duke University Press.
Oyama, Susan. 2001. Terms in tension: What do you do when all the good words are taken? In Oyama *et al*. (eds.), 177–193.
Oyama, Susan, Paul E. Griffiths & Russell D. Gray. 2001. Introduction: What is Developmental Systems Theory? In Oyama *et al*. (eds.), 1–11.
Oyama, Susan, Paul E. Griffiths & Russell D. Gray (eds.). 2001. *Cycles of Contingencies: Developmental Systems and Evolution*. Cambridge, MA: MIT Press.
Pigliucci, Massimo & Gerd B. Müller (eds.). 2010. *Evolution — The Extended Synthesis*. Cambridge, MA: MIT Press.
Pinker, Steven & Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* 95, 201–236.
Pinto-Correia, Clara. 1997. *The Ovary of Eve: Egg and Sperm and Preformation.* Chicago: University of Chicago Press.
Robert, Jason Scott. 2003. Developmental systems theory. In Brian K. Hall & Wendy M. Olson (eds.), *Keywords & Concepts in Evolutionary Developmental Biology*, 94–97. Cambridge, MA: Harvard University Press.
Robert, Jason Scott. 2004. *Embryology, Epigenesis, and Evolution: Taking Development Seriously*. New York: Cambridge University Press.
Robert, Jason Scott, Brian K. Hall & Wendy M. Olson. 2001. Bridging the gap between developmental systems theory and evolutionary developmental biology. *BioEssays* 23, 954–962.
Roe, Shirley A. 1981. *Matter, Life, and Generation*. Cambridge: Cambridge Univer-

sity Press.

Samuels, Bridget, Marc D. Hauser & Cedric Boeckx. In press. Do animals have Universal Grammar? A case study in phonology. In Ian Roberts (ed.), *The Oxford Handbook of Universal Grammar*. New York: Oxford University Press.

Smith, Neil. 1999. *Chomsky: Ideas and Ideals*. Cambridge: Cambridge University Press.

Spencer, John P., Mark S. Blumberg, Bob McMurray, Scott R. Robinson, Larissa K. Samuelson & J. Bruce Tomblin. 2009. Short arms and talking eggs: Why we should no longer abide the nativist–empiricist debate. *Child Development Perspectives* 3, 79–87.

Stich, Stephen. 1975. The idea of innateness. In Stephen Stich (ed.), *Innate Ideas*, 1–22. Berkeley, CA: University of California Press.

Thompson, D'Arcy W. 1917. *On Growth and Form*. Cambridge: Cambridge University Press.

Uriagereka, Juan. 2007. The evolution of language. What songbirds, dancing, and knot-tying can tell us about why we speak. *Seed Magazine*, September 25. [http://www.seedmagazine.com] (last accessed 20 June 2011).

Weber, Bruce W. & David J. Depew. 2001. Developmental systems, Darwinian evolution, and the unity of science. In Oyama *et al*. (eds.), 239–253.

Wexler, Kenneth. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106, 23–79.

Wexler, Kenneth. 2003. Lenneberg's dream: Learning, normal language development, and specific language impairment. In Yonata Levi & Jeannette Schaeffer (eds.), *Language Competence across Populations: Toward a Definition of Specific Language Impairment*, 11–61. Mahwah, NJ: Lawrence Erlbaum.

Wexler, Kenneth. In press. Grammatical computation in the optional infinitive stage. In Jill de Villiers & Thomas Roeper (eds.), *Handbook of Generative Approaches to Language Acquisition*. Berlin: Springer.

Cedric Boeckx
Universitat Autònoma de Barcelona
ICREA & Centre de Lingüística Teòrica
Facultat de Lletres, Edifici B
ES–08193 Bellaterra (Barcelona)
Spain
cedric.boeckx@uab.cat

Víctor M. Longa
Universidad de Santiago de Compostela
Área de Lingüística General
Plaza Isabel la Católica, 2, 2º E
ES–36204 Vigo
Spain
victormanuel.longa@usc.es

★ FORUM ★

The Character of Mind

Wolfram Hinzen, Nirmalangshu Mukherji & Bijoy Boruah

Even if the human species is not the only one that has a mind, it may be the only one that knows that it does. Fascination with our own mind, with what we feel, sense, and think, lies at the heart of human nature and has unsurprisingly become a major part of scientific inquiry itself. Rigorous scientific inquiry into the character of mind has been a part of all major traditions in scientific thought, but the character of these inquiries varied across different traditions, some of which have also been essentially separate for millennia and are only being rediscovered now. Thus, the formal study of grammar was an essential ingredient in the Indian Classical tradition, leading to more than a thousand years of rich and intense discussions in linguistics and philosophy of language in the hands of *Vyakaranvadis* (grammarians) such as Pāṇini, Tolkappiyar, and other authors in their traditions respectively in northern and southern India (Matilal 1990). There is essentially no parallel to this in the Ancient Greek tradition, where not grammar but geometry was the entry point to science. And although Aristotle developed a model of the sentence that has proved relatively stable for two thousand years of linguistic theory (Moro 1997), the first tradition of Universal Grammar in the Western world emerged not before the 1200s in Paris (Covington 2009), where Modistic grammarians viewed grammar as a formatting principle for a species-unique kind of thought. Flourishing across much of Northern Europe by the end of the 13th century, it eclipsed after less than a hundred years when nominalist doctrines entered the scene and logic took pride over grammar again as a meta-theoretic framework. Interestingly, a similar eclipse happened with the grammarian tradition in India as the logico-empiricist framework of the *Nyayaikas* (logicians) became dominant. The next tradition in scientific thinking about human grammar, namely Port Royal, emerged within Cartesian rationalism in the 17th century, and was taken up by Noam Chomsky in the 20th (Chomsky 1966).

With this last tradition we associate the term 'second cognitive revolution', which now is little more than 50 years old. To review it was part of the goals of an international conference convened by Nirmalangshu Mukherji, Wolfram Hinzen,

Funding from the grant 'Un-Cartesian Linguistics' (AHRC/DFG, AH/H50009X/1, to Hinzen), which went into the preparations for the conference and into the writing of this report, is gratefully acknowledged.

biolinguistics

ISSN 1450–3417 *Biolinguistics* 5.3: 274–283, 2011
http://www.biolinguistics.eu

and Bijoy Boruah, on 'The Character of Mind'. It was held at the Indian Institute of Advanced Studies in Shimla (Northern India) from 18–20 March 2011, with the generous financial support of the Institute and the Indian Council of Philosophical Research. Bringing together eminent scholars and scientists from India, Canada, Italy, the UK, and the US — coming from disciplines such as philosophy, psychology, linguistics, law, biology, and physics — the following questions were asked: What has been achieved in half a century of study of the cognitive mind? How does it connect with millennia of human effort to bring light to the structure of our mind, in different traditions with radically different emphases and cultural conditions? Are there lasting insights unifying these traditions? Is the evidential basis clear on which claims about the character of mind can rest?

A look at the history of the science of cognition reveals both essential continuities and discontinuities. The fragility of the enterprise at large impresses itself on the observer: Progress in the study of mind has been very far from linear. Continuities cannot be overlooked, on the other hand. Thus as **Amita Chatterjee**'s (Calcutta/ Kolkata) presentation illustrated, the 20th century debate on whether the representational resources of our mind track a mind-independent external reality, or whether our representational access to the world is rather linguistically mediated, is as well-articulated in the Classical Indian tradition of the *Navya-Naiyāyika*'s (new logicians) as it is in the 20th century Western analytic tradition. A characteristic conflict between the viewpoint of the logicians and the grammarians is to be found in the Indian tradition as noted, as much as it is a leading theme in the downfall of Modistic grammar (Covington 2009), mentioned above, in the turn of 19th century grammarians against the Port Royal tradition of logic (Graffi 2007), or in the turn of 20th century logicians such as Russell against the structure of human language as a valid source of philosophical insight. The persistence of this somewhat uneasy partnership between grammar and logic as possible frameworks for philosophy throughout the course of human scientific rationality marks it as a particularly important focus of further historical and systematic study.

Inquiry into which mental structure is 'innate' to the mind is also a defining feature of scientific rationality as constructed by Plato which, despite numerous attempts to contravene it, is no less alive today, as the contribution of **Susan Carey** (Harvard) on the origin of concepts at the conference exemplified. Interestingly, on the other hand, the epistemology of innateness does not appear to play the same paradigmatic role in the study of grammar in the Indian tradition. Not only is the epistemology different, but also there was a more definite focus on the grammatical mind, as noted, such that grammar was viewed as playing a role in the structuring of both thought and reality. As **Probal Dasgupta** (Kolkata) put it in his contribution, formal grammar in the wake of the Chomskyan framework has 'focused on the grammatical rule as the austere formal object of rigorous statement', thereby ignoring an essential turn that Indian formal linguistics took in the 7th century with the work of Bhartrihari, whose seminal work *Vaakyapadiiya* (On Words and Sentences) inaugurates what we may call a 'substantivist' approach, in which the whole cycle from sentence composition through speaking, hearing, and understanding to fresh composition is identified as the proper object and domain of linguistic inquiry.

As **Godavarish Mishra**'s (Delhi) presentation made clear in this connection, this inquiry into the structure of the cognitive mind naturally points beyond itself even further. While metaphysics has been banned from modern science, it would be naive to conclude that cognitive science has no bearing on deeper and wider questions of a philosophical nature even today. Bhartrihari exemplifies this point, too, when he not only rejects the familiar view of language as mere 'vehicle' of thought (or its 'conveyer-belt', as when language is merely expressive), but maintains that for something to count as knowledge, it has to be given a linguistic form. In this sense, what we know as the world is a creation of language. A world or mind placed outside of language, Bhartrihari maintains, would be inscrutable.

In Bhartrihari's case, these wider epistemological and metaphysical contentions are woven into an intricate fabric of more specific empirical and methodological hypotheses about language that bear on it, such as the question whether meaning is compositional (exhibits a part–whole structure), whether lexical items have an independent meaning, and whether the organization of the word is fundamentally distinct from the organization of the sentence. Dasgupta, arguing for the latter view, exemplifying it with restrictions on recursion in the domain of morphology, was here contradicted by **Anna Maria Di Sciullo** (Montréal), who argued that human language is characterized by a small set of basic operations — a toolkit, including the operation Merge — which can be used to various degrees in different domains, but is implied in morphology as well. Yet the recursions are different, with morphology reflecting arithmetical recursion somewhat more closely than syntax. The question why human language clearly distinguishes these two domains of grammatical organization — the word and the sentence — remains.

While all of these issues are focused around the role of grammar in human cognition, it is abundantly clear that mind is not exhausted by language. In fact, if one takes the Shimla event as any indication, it may be useful to see the human mind as characterized by three broad domains:

(A) (self-)knowledge
(B) grammar
(C) experience

Knowledge, including self-knowledge, has been at the heart of the Western tradition, particularly since Descartes' re-centering of knowledge around the Self and its cogitations, which formed the subject of **Bijoy Boruah**'s (Delhi) presentation discussed below. The same is true of the Indian tradition, such as the *Vedanta* — with an alternative tradition too denying the Self, such as the *Carvakas* ('materialists') and the Buddhists. Grammar is clearly central to knowledge, for it appears that we can obtain no know-ledge in the relevant sense in domains of cognition where the relevant structural formats don't exist: associative cognition, say, or emotional cognition. One of the most crucial questions in the study of cognition thus is what difference grammar, which is species-unique, makes to the character of mind, and perhaps even to the origin of our species. Yet, as **Ned Block** (New York) put it in his presentation, reason is not the essence of the mental; or, to put this around, the mental — in the sense of awareness of phenomenal experience — is not the essence of reason. Hence, there is a distinct and crucially different third domain of experience, which confronts us with explanatory problems distinct from either those posed by knowledge or by gram-mar.

Block's topic — phenomenal richness of experience, which is often unlike what we take it or report it to be — illustrated how fascinating and difficult the question becomes what we really experience when language is not there to structure that experience and to report it, or when such reports are seen always to be a function, not only of conscious experience, but cognitive and affective responses to task-specific demands as well. Somewhat similar problems keep arising in comparative cognition. As Block notes, 'even our interpretations of animal research must ultimately be based on human first person reports'. Evidence for the richness of the non-linguistic animal mind is rich and undeniable, yet what it is exactly that an animal is thinking at a particular moment (which concept, in which structural arrangement), has proved to be an elusive question. Is an animal thinking at all if it has no concept of a thought (Davidson 1984)? If it has no words, are its 'concepts' like the meanings that words have when they occur in sentences (where, in particular, they all have a grammatical *category*, which no concepts as such have)? These questions came up in discussions on several papers in the conference.

When language gives out, in short (and by consequence there is no space of shared words linked to entities in the external world that we can use to tap into a hearer's mind), it may be that reality becomes harder to delineate. Ingenious methods now have to be devised to get a glimpse of this non-grammatical reality; neuroimaging, say. Yet the continuous neural activity imaged with today's technologies proves not to be of the right grain to capture experience where it has a content that is discrete. The situation gets more complicated still when we realize, as **Barry Smith** (London) argued, human 'conscious experience' itself — in traditional terms, the experience of a Kantian or Cartesian 'subject' — is not the unified phenomenon it has long been taken to be in at least the Western modern philosophical tradition. A look at 'abnormal' experience in mental illness after brain damage reveals that a 'normal' subject's experience may also not be as normal as we thought. Hidden beneath something as simple, familiar and basic as a feeling of 'agency' lie myriad interacting systems in the brain that sustain the illusion of unity where none exists. While some of the 'abnormal' experiences are traceable to aspects of language impairment, some are not. As with Block, Smith left it open as to how much of the structure of human experience can be traced to nonhuman organisms.

This essential *fragility* of the construct we call the 'Self' is a notable new topic for systematic inquiry. In contrast, Bijoy Boruah's talk served to remind participants of the undeniable intuitive force of traditional Cartesian intuitions on the ultimate simplicity and unity of the way in which the Self is presented to itself. Any way of objectifying the content of the experience of the Self indicates that we have missed the target of our inquiry, the subject. Here we enter a world of reflection of the human mind on itself that is more structured, and in particular strictly distinguishes the 'I' from any 'you' (including the you into which the I turns itself when addressing or scrutinizing itself), and both of these from any 'it'. For this move, relevant structural resources are required which may well be grammatical: It it is not clear whether any such form of self-reference can be sustained in the absence of a system of grammatical person, which we know plays a crucial role in the convergence of syntactic derivations (see e.g. Sigurðsson 2008, Longobardi 2010). Also, we need to explain that much of the grammaticality of 'I' and 'you', such as agreement structures, go through without assuming self-reference; otherwise, it will be impossible to deny the Self, à la David Hume or the Buddhists.

It appears, then, that some cognition is pre-grammatical, existing in pre-verbal infants and non-human animals, some is post-grammatical, or at least stands in some inherent relation to grammar. As Carey illustrated, before human beings create scientific theories, mathematics, literature, moral systems, and complex technology, all of which are culturally constructed and require grammar among other things, there is a rich world of concepts characterized by inferential roles and representational functions. How they are structured — given that they are not structured grammatically — is an open question. In the case of a concept such as number, the intrusion of grammar and a public language in the development of the infant may boost innate representational resources in a way that is not innately pre-specified. The richer a pre-linguistic world of concepts becomes, the more astounding and mystifying is the transition from pre-human hominins to our early African ancestors. Why, if the conceptual world of the Neanderthal and its immediate predecessors (let alone the chimpanzee) is so rich, is their culture so poor, and so un-suggestive of the innovation and creativity that marks a modern human culture in the Aurignacian as much as it marks the language faculty that supports it? And what is the nature of the major transition in infant development that happens around the 4th year of age, as Carey argued for with a range of examples?

As Carey suggests, it may make sense to think of the matter in terms of a hierarchy of increasingly abstract representations, which may set out with percepts, continues with concepts and their inferential roles, and at some stage includes images. This was the topic of **Mohan Matthen**'s (Toronto) presentation, who attributed propositional content as well as 'force' to such imagic representations (since one can anticipate, recollect, expect them, for example), thereby foreshadowing formats of cognition normally reserved for an illocutionary format of representation, i.e. the linguistic case. The entire hierarchy seems to obtain even before there are any *words*. At this junction a crucial question arises: What difference do words actually make?

Addressing this very question, **Wolfram Hinzen** (Durham) noted that the move from concepts to words marks a difference in grammaticality — every word has a grammatical category — which in turn accounts for the fact that words can occur in sentences: They are parts of speech. Importantly, most words with a substantive lexical content can be re-categorized, moreover: For example, the root concept KILL can be grammaticalized as a noun (*a quick kill*) or as a verb (*kill Bill*), with mixed forms in between, such as nominalizations of the verbal form (*the killing of Bill*). These occurrences of the same lexical concept KILL clearly have nothing to do with a difference in semantic content, which remains the same throughout. What differs, rather, is ontology: whether this concept is referred to as an object (nominal case), as an event (verbal case), or as both. The external world, however, has little if anything to do with these differences in ontology: It could be exactly the same, and yet a speaker will refer to it with the nominal form on one occasion and the verbal one on another. The difference, thus, is a difference, not in semantic content, but deixis: The way of referring (the *modus significandi*, in traditional terminology). The relevant forms of deixis are inherently grammatical, moreover, and thus not to be found in non-grammatical beings. Grammar, in short, is a device of extended deixis: We use language to point hearers to objects, facts, and truths, which no creature that merely has concepts can do.

Hinzen's take reflects a certain departure from the viewpoint of grammar as a purely formal object, even though the formality of generative grammar in the past has reflected a methodological decision, rather than an empirical claim, about the substantive nature of the object under study. Nonetheless, the formal treatment of the computational system underlying language is a crucial move within Minimalism, when it attempts to see the computations in the language faculty as an instance of computations in a wide range of species, or indeed in physical nature as such. Di Sciullo's presentation precisely raised this question: How far the most fundamental computational operations reflect generic processes in nature that can be found in, say, cell division or organic growth as well. Universal constraints on linguistic computations, too, may be generic in the sense of reducing derivational complexity or avoiding the number of choices in a derivation. If so, a central question for the linguist is which language-specific operations need to be added to the fundamental principles of computation and recursion. Neurophysiological experiments suggest that the brain is specifically sensitive to crucial asymmetries arising derivationally, like between complements and non-complements or between mono- and bi-phasal structures. Even at this level of specificity the question arises whether the constraints in question are generic in nature, or else linguistically specific.

How much, then, is the grammatical mind really part of the physical world, as opposed to a joint of nature that cannot be conceptualized in other than grammatical terms? The question arises if one accepts, with Katz & Pesetsky (2009) and Mukherji (2010), that the joint of nature includes more than language, namely music and arithmetic as well. This peculiar triad and related domains may represent a unique configuration in nature that is simply not found anywhere else than in the grammatical mind itself, where they are used to compute sound and meanings. Looking for generic operations in the language faculty and regarding the latter as arising from biological processes not specific to the human mind is a well-motivated recent path which resists this conclusion. Yet, as **Nirmalangshu Mukherji** (Delhi) argued, the conclusion may nonetheless be right. To put the conclusion differently, talk of 'computational systems' outside of the human species — as when desert ants and foraging bees are said to have it when computing paths of motion— may be a move guilty of equivocation in the very term 'computational system'. As is worth noting in this regard, the best evidence for relevant computations does not come from the non-human primate lineage, which forms the most relevant comparison class: Chimpanzees don't vocalize, and their thought system appears to be radically different from ours. The grammaticalization of sound and meaning may thus — consistent with Hinzen's story — *create* the very meanings that sentences encode and the very sounds that externalize them. Outside of a grammaticalized world, they would simply not be found, and where a computational system in the ant or bee brain has been posited, either a more biological story or a specific non-symbolic story (Bickerton 2009) may have to be sought that makes sense of the data.

Coming back to our earlier thoughts about 'hierarchies', it is clear that when we have moved from percepts to concepts to images, and from there on to words and sentences, we are nowhere near the end of the hierarchy of mental complexity. Thus, while a moral mind is surely necessarily a linguistic one, the naturalistic analysis of its grammaticality tells us relatively little about its moral content. As **John Mikhail** (Washington) discussed with reference to a rich tradition of inquiry lasting several centuries, however, the moral mind is nonetheless crucially a generative one as well: A moral being is capable to compute, on the spot, a potential infinity of complex moral judgements appropriate to an occasion, whose perceptual and physical feature will typically radically underdetermine the judgements in question. The rationality of these judgements is furthermore clearly not rational in the sense of consciously rationalizable by the subject in question, creating an analogy with a major insight in regards to the grammatical mind associated with the second cognitive revolution in the 1950s. The generative principles of moral judgment may thus be as inaccessible to conscious introspection as the principles of grammaticality. Yet, as was discussed at length, differences between the moral and the linguistic faculty nonetheless abound, with the former for example being subject to learning, instruction, and moral conflict in a way that linguistic judgements are not. Morality may also not allow for the methodology of individualism, in the way that grammar has at least been thought to do (though Hinzen's story in regards to the deictic significance of grammar suggests reasons for skepticism in this regard). An account of the moral mind that appropriately identifies both the overlap and the differences between the two kinds of computations needs to predict these differences, so that talk of a 'moral grammar' in the brain is able to avoid the danger of involving a metaphorical extension of the term 'grammar' — much as, on Mukherji's account, talk about grammar in ants and bees may involve such metaphorical extensions.

Difficulties with an understanding the moral mind in grammatical terms again illustrate conceptual obstacles when attempts are made to transcend the naturalistic study of the grammatical mind as pioneered by Chomsky half a century ago. As noted above, opening up cognitive science to the realm of the phenomenal and the Self moves us well to the boundaries of scientific inquiry, and perhaps in part beyond. Yet, it is noteworthy that the scientific study of consciousness has been burgeoning for many years, and much insight has been obtained. Where the moral mind is our topic, on the other hand, naturalistic inquiry will now confront normative issues that the generative approach to grammar has sought and managed to avoid.

That said, it is surprising which aspects of the mind this approach has now succeeded to illuminate. As **Giuseppe Longobardi** (Trieste) illustrated, the history and distribution among these is an excellent example, to an extent that the study of the history of human languages becomes a domain of inquiry from which to obtain a novel argument in favor of a computational approach to the mind in the sense of the generative program and its study of language from a mentalist point of view. Reconstructing linguistic phylogenies has until recently operated at a relatively 'surface' (or phenotypic) level of linguistic description, often focusing on words and their histories or relatively superficial structural patterns, at the expense of the 'I-linguistic' mechanisms studied for many decades in generative linguistics. As such it has certainly not been able to take us beyond the threshold of 10,000 years of human history. Yet, even before that date, major human and linguistic diversity must have existed. As Longobardi explained and illustrated, just as genetics has introduced genetic and molecular markers (cf. Cavalli Sforza *et al*. 1994) which are more abstract and only indirectly connected to external phenotypical traits, historical linguistics can now make a similar move using the resources of parametric analyses of grammatical diversity, leading to more stable and reliable historical indicators of phylogenies, with a potential to reach further back into the human past (Longobardi & Guardiano 2011). None of this is in any conflict with the fact that grammatical structure reflects cultural history to some extent (Dunn *et al*. 2011).

Much of the discussion thus indicated the need for a closer study of origins of human language. Specifically, was there a relatively recent speciation event that definitively separated the humans from the rest of the post-chimpanzee hominid line to lead to the emergence of language and its wide effects on human cognition? **Timothy Crow** (Oxford) could not attend the conference unfortunately. But his work, including the extended abstract submitted for the conference, was frequently mentioned. Given that human cognitive capacities far exceed those of our primate relatives, Crow asks, if the transition was saltational what was the mechanism? Following the characteristic asymmetry of the human brain, Crow (2010) notes that the human brain is four-chambered (right and left, anterior and posterior) and circuitous with respect to heteromodal association cortex by contrast with the bilateral equality (anterior and posterior) of the chambers of the generalized mammalian brain. This suggests that a discrete speciation event took place about 160KYA, that perhaps the ProtocadherinXY gene pair was involved, and that the effect was to render the human brain 4-chambered with respect to heteromodal association cortex. From this arose the capacity for language. According to Crow, the compartments of the human mind are identified sequentially with speech perception, meaning, thought and speech production, or more technically, with perceptual, conceptual, intentional and articulatory capacities. It is interesting that this is exactly how the 'external systems' of language, the sensorimotor and the conceptual-intentional systems, are conceptualized in biolinguistics. This still leaves open the crucial issue of how the combinatorial system of Merge itself emerged.

Overall, it is thus clear that the study of the grammatical mind over the last half-century has raised deep issues in regards to both the unity and the diversity of the human mind. It has not only pointed to the epistemological significance of grammar but also to other inquiries into the cognitive mind that are at the frontier of inquiry today. Very clearly, the issues of linguistic theory point beyond the empirical properties of human languages, to the origins of our species and of human variation as such.

The conference concluded with a talk by the physicist **Partha Ghose** (Kolkata) focusing on the famous discussion between the poet–artist–philosopher Rabindranath Tagore and the physicist Albert Einstein on the character of scientific truth (Marianoff 1930). While Tagore held that all truths, including truths of physics, can only be human truths, Einstein urged that physics will be impossible unless we entertain an external reality independent of the human mind. Ghose suggested that this classic realism/antirealism debate is also reflected in two apparently conflicting directions in contemporary cognitive science. According to Ghose, proponents of 'embodied cognition' such as Fransisco Varela (Varela *et al*. 1992) hold a view closer to Tagore, while formalists/computationalists such as Chomsky perhaps hold an Einsteinian view. Ghose held that the measurement problem in quantum theory is a test case for physics. If quantum theory is a general theory of the Universe and the measurement problem its inevitable consequence, then even quantum theory could be viewed as 'embodied' in the sense that it necessarily incorporates the effects of human perception. This is precisely the reason why Einstein denied that quantum theory is a 'complete' theory. The issue obviously touches the very heart of cognitive science, including biolinguistics, since cognitive science attempts to use the human mind to study itself.

References

Bickerton, Derek. 2009. *Adam's Tongue: How Humans Made Language, How Language Made Humans*. New York: Hill and Wang.

Cavalli Sforza, L. Luca, Paolo Menozzi & Alberto Piazza. 1994. *History and Geography of Human Genes*. Princeton, NJ: Princeton University Press.

Chomsky, Noam. 1966. *Cartesian Linguistics*. New York: Harper and Row.

Covington, Michael A. 2009. *Syntactic Theory in the High Middle Ages*. Cambridge: Cambridge University Press.

Crow, Timothy J. 2010. The nuclear symptoms of schizophrenia reveal the four quadrant structure of language and its deictic frame. *Journal of Neurolinguistics* 23, 1–9.

Davidson, Donald. 1984. Thought and talk. In Donald Davidson (ed.), *Inquiries into Truth and Interpretation*, 155–170. Oxford: Clarendon Press.

Dunn, Michael, Simon J. Greenhill, Stephen C. Levinson & Russell D. Gray. 2011. Evolved structure of language shows lineage-specific trends in word-order universals. *Nature* 473, 79–82.

Graffi, Giorgio. 2007. *200 Years of Syntax: A Critical Survey*. Amsterdam: John Benjamins.

Katz, Jonah & David Pesetsky 2009. The identity thesis for language and music. Ms., Cambridge, MA: MIT. [http://ling.auf.net/lingBuzz/000959]

Longobardi, Giuseppe. 2010. Reference to individuals, person, and the variety of mapping parameters. In Henrik Høeg Müller & Alex Klinge (eds.), *Essays on Nominal Determination: From Morphology to Discourse Management*, 189–211. Amsterdam: John Benjamins.

Longobardi, Giuseppe & Cristina Guardiano. 2011. The biolinguistic program and historical reconstruction. In Anna Maria Di Sciullo & Cedric Boeckx (eds.), *The Biolinguistic Enterprise*, 266–304. Oxford: Oxford University Press.

Marianoff, Dmitri. 1930. Einstein and Tagore plumb the truth: Scientist and poet exchange thoughts on the possibility of its existence without relation to humanity. *New York Times*, August 10. [http://www.mukto-mona.com/Articles/einstein_tagore.htm]

Matilal, Bimal Krishna. 1990. *The Word and the World*. Delhi: Motilal Banarasidas.

Moro, Andrea. 1997. *The Raising of Predicates*. Cambridge: Cambridge University Press.

Mukherji, Nirmalangshu. 2010. *The Primacy of Grammar*. Cambridge, MA: MIT Press.

Sigurðsson, Halldór Ármann. 2004. The syntax of person, tense, and speech features. *Italian Journal of Linguistics / Rivista di Linguistica* 16, 219–251.

Varela, Francisco J., Evan T. Thompson & Eleanor Rosch. 1992. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.

Wolfram Hinzen
Durham University
Department of Philosophy
50 Old Elvet
Durham DH1 3HN
United Kingdom
wolfram.hinzen@durham.ac.uk

Nirmalangshu Mukherji
University of Delhi
Department of Philosophy
Probyn Road
Delhi 110 007
India
somanshu@bol.net.in

Bijoy Boruah
Indian Institute of Technology
Faculty of Humanities and Social Sciences
Hauz Khas
New Delhi 110 016
India
boruah@iitk.ac.in

Biolinguistics

Volume 5, Issue 3

Summer 2011

TABLE OF CONTENTS