CHAPTER 3

'Biolinguistics'

Some foundational problems

Robert Levine Ohio State University

The 'Biolinguistics' program seeks to establish specific neuroanatomical models corresponding to the representations and operations characterizing the species-specific language faculty in human beings. Yet after decades of research, no neural structures corresponding to specific linguistic structures, rules, constraints or principles have ever been identified. A key to biolinguistics' failure is, I suggest, its long-term adherence to two dubious assumptions: (i) a kind of literalism in envisaging the relationship between neural anatomy and linguistic representations, reflecting a seriously misconstrual of Marr's (1982) tripartite division of cognition, and (ii) a view of such representations as objects fundamentally different from other components of human cognitive capacity. (ii) rests on the premise that phrase markers are the optimal formal representation of natural language sentences, despite major empirical difficulties that syntactic accounts based hierarchical phrase structure face in handling a wide variety of grammatical patterns, including non-canonical coordinations and ellipsis constructions. In contrast, proof-theoretic approaches such as type-logical grammar do not face these difficulties, and their foundational assumptions link language to the higher-order cognitive functions supporting deductive reasoning. This conclusion suggests a promising alternative to the current, essentially result-free 'Biolinguistic' paradigm.

Keywords: biolinguistics; linguistic ontology; logic; syntax; semantics

1. Preview: What's wrong with 'biolinguistics'

The appearance of scare quotes in my title and in the heading to this introductory section points to a critical distinction maintained throughout the discussion below. There is a perfectly respectable interpretation possible for the term 'biolinguistics', which came into common use during the past decade, in which the term identifies a domain of inquiry which aspires to identify neurological (and other anatomical) structures corresponding to the human linguistic capacity. This aspirational use of the term seems quite reasonable, in the same way that 'exobiology', denoting the study of extraterrestrial life, is. There are actually no exobiological results - we have yet to detect life anywhere except on Earth; but there is a very well developed body of interdisciplinary studies about the conditions under which life can evolve, where and how in the universe such conditions might be realized, what form it might take, and so on. In the same way, there are, as I argue below, no results that would come under the heading of a biocomputational theory of linguistic cognition that bear comparison with, for example, proposals that would legitimately describable as the biocomputational theory of visual cognition. But, rhetorical posturings aside, no one, regardless of theoretical framework or ontological stance, denies that the human capacity for language rests on a neurological basis. The evidence for this conclusion is overwhelming and has been taken as a given from the time Broca's discoveries became widely known. Several of Brodmann's areas in the neocortex are heavily implicated in speech and language comprehension and production areas. What remains a complete mystery is how specific neuroanatomical structures 'run' native speakers' knowledge of their languages, and, most contentiously, what the relationship is between that knowledge and the formal contents of the various grammar architectures proposed by linguists as the optimal theory of those languages. A field called biolinguistics which carefully attempts to lay out the many issues, lines of investigation, and potential problems with such investigation surely has a claim to scientific credibility, even in the absence of any robustly confirmed proposals. But that isn't the sense of the 'biolinguistics' referred to in the header, whose scare quotes are deliberate.

What I'm referring to is rather the sense of the word as it appears in the following comment by Gillian Ramchand, quoted in Martins and Boeckx (2016: 2).

> There are parts of generative grammar that I do not feel a particularly strong part of. For example, I am not sympathetic to recent trends in Biolinguistics, which to my mind is guilty of extreme Overreach in attempting to connect linguistics to Biology. I think it gives the whole field a bad name. The granularity gap and the terminology gap (to put it in Poeppel's terms) are still too great to sustain the specific kinds of proposals that are being taken seriously in this sub-group.

The kind of idea that Ramchand appears to be objecting to is the proposition, seriously offered by prominent long-time adherents of P&P approaches, that grammars are *literally* instantiated in neural tissue. An extreme example is the statement, made in the course a debate with Dan Everett on the relationship between linguistic capability and human biology, in which Anderson and © JOHN BENJAMINS PUBLISHING

the view that if linguistics were as we claim, syntactic trees should be visible in CAT scans, asserted in Everett (2001), seems to reflect more on the adequacy of current brain imaging techniques than it does on the nature of language.

(Anderson & Lightfoot 2006: 81)

On the face of it, this observation is analogous to a prediction that with sufficiently powerful telescopes capable of inspecting regions of space subject to enormous curvature due to to some compact massive object, we should be able to visually detect the field equations of general relativity, and has about as much credibility. But views of this sort are sufficiently common that Ramchand identifies them as trends. The authors speculate that possibly she is refering to "biolinguistic' proposals made [by] those who use term as just another name for the kind of linguistic theorizing that has been common practice for decades" (Martins & Boeckx 2016: 2), but it strikes me as far more likely that her remarks were directed at precisely the kind of view expressed by Anderson and Lightfoot's retort to Everett, identified in her statement as a common enough view of this field to be a source of intellectual discomfort with its name. Nonetheless, Martins and Boeckx's suggestion calls attention to an important thread in the development of that view.

The story here runs along the following lines: for decades, the standard view of natural language grammars is that they are formal accounts of that capacity - or, more precisely, of the specific instantiation of that capacity in the minds of adult native speakers. On this view, dominant in transformationalist circles since the early 1980s, such speakers have fixed various parameters left open in a panhuman 'language organ', an initial state of the brain which, exposed to linguistic data in infancy and childhood, sets values for these parameters that successive approximate the input data until a stable state, the adult grammar, is achieved. This model of the language acquisition process, in which the initial state consists of a set of general principles with the values of certain parameters undetermined at the outset, and only set in accordance with input data during the linguistic maturation process, has been the dominant paradigm in the field for nearly four decades more than half of the entire career of generative grammar, supposedly accounting for both the course of language development in individuals and the range of possible variation among natural languages.¹ In earlier work, Chomsky identified

Remarkably, however, there is no actual theory of the parameters which play such a crucial role in this 'Principles and Parameters' (P&P) view, nor any concensus on critical questions such as what dependencies hold amongst the set of parameters, or even what parameters there are in the first place. See Newmeyer (2005) for a broad and deep critique of the network of assumptions, and their confrontation with results from typological research. © JOHN BENJAMINS PUBL

linguistic cognition in what seem to be metaphorical terms, seemingly hedged to avoid any specific ontological commitment to the nature of this mental ability:

> We may regard the language capacity virtually as we would a physical organ of the body and can investigate the principles of its organization, functioning and development in the individual and the species.

> > (Chomsky 1976: 46, emphasis added)

Somewhat later, however, the 'stable state' corresponding to adult grammars is explicitly identified as a having a literal physical being:

> a mentally represented grammar and UG are real objects, part of the physical world, where we understand mental states and representations to be physically encoded in some manner. Statements about particular grammars or about UG are true or false statements about steady states attained or the initial state (assumed fixed for the species), each of which is a definite real-world object, situated in space-time and entering into causal relations.

> > (Chomsky 1983: 156–157; emphases added)

Linguistic capability now appears to be regarded, not 'virtually' but literally, as a physical object, with speakers' adult language capacity realized as a specific set of neural structures.

From this point, it is conceptually only a rather short step to the reductio that grammars themselves, the formal expression of that capability, should, with sufficiently sophisticated technology, be identifiable in the neural scans such technology would make available, as in Anderson and Lightfoot (2006: 81). Postal (2009), commenting on the passage from their reply to Everett cited above, observes that it makes the elementary category error conflating tokens (concrete objects) with types (abstractions), but the problem is still more pernicious. A phrase structure tree is simply a graphical representation of a set of nodes under two mutually exclusive partial orderings. The advantage of such a representation is legibility: by assigning the vertical representations to the domination partial ordering, and the horizontal to the linear precedence ordering (itself a convenient graphic encoding of a temporal, not spatial, order), we avoid the delimiter-counting problem which makes labeled bracket notation so hard to parse visually. But there is nothing more intrinsically tree-like than bracket-like about the set-theoretic objects we use these geometric or typographic conventions to illustrate, and there are any number of alternative encodings for these objects, e.g., the attribute-value matrices commonly employed in Head-driven Phrase Structure Grammar, Lexical-Functional Grammar, Functional Unification Grammar and others. Anderson and Lightfoot are apparently under the impression that a sufficient increase in the sophistication of our imaging technology would see the neural enscription of dominance and precedence relations in ANY particular conventional notation. The key problem with © JOHN BENJAMINS PUBL

the view just cited, then, is the implausibility of the conviction that neural arrays - the kind of thing that brain scans reveal - can literally embody set-theoretic, algebraic or category-theoretic objects and mathematical relations. As I discuss below, neural function can reflect a relatively concrete relation such as amplitude, and this is the basis for the success of the Marr framework for mathematically modeling the visual system. Marr's position was - as will become evident - far closer to Everett's than to Anderson and Lightfoot's.

The first problem with this fundamentalist or, as I'll call it, literalist interpretation of the biolinguistic perspective, as exemplified explicitly in Chomsky (1983) and Anderson and Lightfoot (2006), is thus the kind of category error noted by Katz (1981), Katz and Postal (1991), Postal (2009), Behme (2012), Behme (2013), inter alia, and at least implicit in much of recent and current work adopting this perspective. Advocates of the neurological reality of syntactic representations (including Anderson & Lightfoot 2002 themselves) have, however, often invoked psychophysical models of other cognitive domains to justify this premise of the biolinguistic framework, in particular the watershed research of Marr and his associates in the 'Visionaries' group studying the human visual system at MIT's Computer Science and Artificial Intelligence Laboratory. In § 2 below, I outline the critical contribution of Marr's work on edge detection - arguably the foundation of his detailed mathematical formalization of visual perception - and suggest that Marr's actual accomplishment offers no support at all to the claims made in the biolinguistic literature along the lines in Anderson and Lightfoot; on the contrary, he explicitly rejects a view very close to what they themselves assert in the above quotation.

Neurological literalism is not, however, the only questionable aspect of the biolinguistic framework which needs to be critically scrutinized. The passages quoted above from Chomsky, particularly the second one, contain a claim of socalled 'domain specificity', implicit in the reference to the anatomical composition of the human body. Invocation of physical organs carries with it a strong implication resting on the dedicated nature of such organs. The cells of the kidney, brain and liver are fundamentally different in structure from each other and from any other kind of cell type; they evolved to subserve a specific range of functions within the organism, and the arrangement of these cells into the tissue configurations each organ comprises is unique, corresponding to the unique work that that organ carries out in maintaining homeostasis in the body. The analogy to linguistic ability seems to be a conclusion drawn from the supposed disconnection between the critical predicates appealed to in generative grammar at various stages ('command', 'c-command', 'max-command', 'movement rule', 'cyclic node', 'governing category', 'barrier' etc.), on the one hand, and other cognitive/sensory modalities on the other. Such notions have played a crucial role in syntactic accounts of © JOHN BENJAMINS PUBL

islandhood patterns and configurational conditions on anaphora (under the misnomer 'binding theory'). More recently however, both islandhood and conditions on anaphora have been argued - with extremely strong support in especially in the case of the former - to originate in a mix of processing, pragmatic and prosodic factors. These points become critical in the discussion in § 3. Domain specificity is perhaps a more generally held postion, at least publically, than neural literalism, but it has no better a base of empirical support.

Both aspects of the 'biolinguistic' speculation thus turn out to be logical consequences of long-held generative assumptions. The 'literal instantiation' aspect is in effect the reductio of the position that grammars are themselves the form of human knowledge of grammar. In itself, a grammar could serve as a formal ultimately mathematical - model of that knowledge, without any particular ontological commitment (apart from one's view of the reality of mathematical objects), a position completely compatible with a Platonist characterization of grammar. Chomsky has however consistently rejected any position but the naturalist interpretation of NL grammars: the latter constitute the literal form of the 'implicit knowledge' native speakers have of their language. And domain specificity, ultimately, is rooted in the same source: if NL grammars are the content of human knowledge of language, and that content is inscribed in neural tissue, then not only does neurological literalism follow straightforwardly, but, to the extent that the content of grammars looks nothing like any other aspect of human cognition, so does domain specificity. But as I argue directly, both positions are quite vulnerable, resting as they do on a variety of a priori arguments whose empirical bases are fragile at best, and are interlinked in a way which augurs poorly for the future of the aspirations of 'biolinguistics' (as vs. much recent and current research on the neuroanatomical basis of linguistic ability, some of which is discussed below in § 2 and § 4) to one day become a genuine domain of knowledge.

Visual cognition: The role of early edge detection 2.

By far the most empirically successful formal model of a cognitive module arose out of the research paradigm inaugurated by David Marr and his students and colleagues had a transformative effect on vision research in particular, and the standards and practices of the emerging domain of cognitive science and the diverse disciplines it comprised, of a scale at least as great as that of Richard Montague in semantics. The most comprehensive formulation of not just Marr's specific scientific discoveries, but the methodological architecture in whose terms he reformulated the agenda of research in cognition, is given in Marr (1982). For Marr, even more critical than any specific proposal that he offered, was the decomposition of © JOHN BENJAMINS PUBL

the scientific investigation of sensory modalities into four separate phenomenological domains, with concommitant strategies for giving an adequate account in each respectively:²

- The physical problem: Define the cognitive objective as the solving of a physical problem in space and time.
- The mathematical task: Model the physical problem as a specific mathematical task to be accomplished, where, ideally, components of the task will prove to correspond to components of the cognitive operation under study.
- The algorithm design: Identify plausible candidates for the procedure that implements the mathematical computation.
- The wetware components: Hypothesize an explicit neural circuitry which is optimized to instantiate the candidate algorithm(s) under investigation.

Using this division of labor, Marr offered mathematical simulations of the human visual system, and its embodiment in specific neural complexes, that underwrote every aspect of the 'mental organ' trope ever used to defend that characterization of human linguistic knowledge, and, unsurprisingly, for much of the late 1970s through the 1980s, the work of Marr and his colleagues' work was explicitly invoked by Chomsky and his circle as the exemplar for cognitive research, paralleled by generative grammar as a theory of the 'mental organ' for language. But Marr's paradigm is radically antithetical to the central neurological literalist claim, and there is, as I argue below, not a single currently known fact about the neural basis of language which gives even a hint of support to this claim.

The physical problem and its mathematical formulation 2.1

2.1.1 Generalities

To provide an explicit theory of how the visual system yields detailed mental scene descriptions based on the stimulation of receptor cells in the retinas, it is necessary to identify the basic operation(s) required to jump-start the process of constructing such scene descriptons. From Marr's perspective, these operations correspond to a parsing problem: interpreting the intake of the field into a collection of combinable primitives with specifically visual content which will be assembled into higher-order representations containing all the information human sight delivers; his answer to this primary question - where does vision start? - was that we perceive distinct objects in our visual field; i.e., vision begins with the detection of

MPANY A very similar breakdown of phenomenological levels is outlined in Soames (1984) © JOHN BENJAMINS PUBL

the abrupt discontinuities in spatial properties that we identify as edges. Physical edges will correspond in the visual field to changes in intensity, and so the first task for the visual system, and therefore for any successful mathematical model of that system, is the detection of local discontinuities at all scales while supressing finegrained information corresponding to noise in the optical signal.

This general objective has two components. In the first component, the raw information about levels of light intensity provided by neural arrays in the retina must be blurred, so that small-scale fluctuations are buried. A standard mathematical routine for carrying out this smudging of the intensity information provided by the retina is the use of CONVOLUTION, a familiar technique from signal processing engineering described below. The second component is the identification of the prominent discontinuities in the blurred image, which will be interpreted visually as the edges of discrete objects in the visual field. The key question is what such edges should correspond to mathematically, and the heart of Marr's solution was the identification of a differential operator which, applied to the array of intensity values representing the blurred visual field, would correctly pick out sets of points in that field corresponding to a consistent discontinuity. We outline in turn these two aspects of Marr's solution.

2.1.2 Smoothing out the visual field

The physical channel for the visual field begins with photoreceptor cells in the retina, whose central foveal region comprises the networks of cells providing maximum resolution. While there are on the order of between one and two hundred million receptor cells in the whole retina, the number of foveal cells is much smaller, by a factor of approximately a thousand. Simple photoreceptors (rod and cone cells) are activated by light in the visible range, and in turn pass this information up to successively higher intermediate cell layers. A significant amount of data processing has to occur along the way: the information carried by the outermost level of photoreceptors is in effect nothing but a large matrix of pixels (often referred to as a grey-level array) each of which corresponds to a number representing the intensity of the light falling on that microscopic portion of the retina - a far cry from the detailed scene descriptions involving texture, reflectance properties and relative distance that correspond to our immediate visual experience. In Marr's view, the critical point of departure in translating what happens in our retinas into what happens in our visual awareness had to be an account of the fact that we perceive *distinct* objects in our visual field, i.e., vision begins with the detection of boundaries. The first question, then, is how the brain is able to identify changes in intensity which are prominent enough to be diagnostic for a real boundary in physical space, and which are CONSISTENT © JOHN BENJAMINS PUBLISHING

This turns out to be a far from trivial task. Like all other physical systems, what Aristotle called the sensorium of sight is a noisy channel. Local fluctuations in the behavior of the biochemical medium through which neural impulses flow seriously obscure the basic information reflecting the location and properties of objects in space. Adjacent pixels in grey-level arrays may differ markedly in values, and the initial challenge for any model of the visual system is to explain how the mind distinguishes random discrepancies from the systemic differences in value that correspond to the boundaries of objects in three-dimensional space. One of Marr's signal contributions was his insistence that to solve this essentially neuropsychological puzzle, it was first necessary to construct a predictively successful mathematical model of this information-processing task - in his view, a correct characterization of a kind of smoothing operation, leveling out local spikes and troughs in the visual signal, translating the initial array into one in which the largest-scale differences remain. It is these surviving discrepancies which would then be the basis for inferences about the presence of edges in the physical environment.

Defining the problem of edge detection as, in part, the blurring out of differences that do not carry relevant information allowed Marr to treat the extraction of edge information via a number of mathematical tools - e.g., so-called Gaussian smoothing - familiar from image processing techniques pioneered in the late 1950s (in particular Leipnik 1960 and unpublished work of Irwin Sobel and Gary Feldman at the Stanford Artificial Intelligence Laboratory). The key idea behind all such blurring technologies is to replace the numbers in the original input - the grey-level array, in the case of visual processing - with a new array of numbers which eliminate all but the largest discrepancies at any given scale of resolution. For visual cognition, these discrepancies are robust clues to the presence of edges, but everything hinges on the details - in particular, exactly what method one uses to modify the raw input data of the grey-level array.

One reliable approach to this problem is based on the premise that a physical edge will correspond in the visual field to something like a step: essentially the same light intensities will be measured over all points in a certain neighborhood up till the edge, where a major change will appear in the intensity value at all points along the edge. Suppose we determine the value of the blurred grey-level array at each point in the visual field by, very roughly speaking, averaging the values of the surrounding points along with the value at the point in question. Since the desideratum for any two adjacent points is to have very close to the same intensity value unless there is a good (physical) reason for them not to, a sensible way to proceed is to *weight* the numbers that go into the average for a given pixel P, so that the contribution of pixels to the average drops off smoothly the further they are from P. For two pixels X_1, X_2 , which are immediately adjacent to each other, these © JOHN BENJAMINS PUBL

'distant' values will obviously be very nearly the same for X_1 as for X_2 so that unless the difference in value of the two pixel points themselves is large, the modified values at the two points will wind up being quite close. For a variety of reasons, the optimal 'weighting' function which determines the modified grey-level value at any point is the so-called 'bell curve' function, the Gaussian: one multiplies the value at a given pixel-point by the highest value of the Gaussian, while closely neighboring points at the same distance are multiplied by a slightly smaller value, and so on until, at a certain distance, the contributions have dwindled down to essentially negligible values. The revised grey-level value at a given pixel-point is then obtained by adding all the contributions together as calculated from that point. In effect, differences in intensity levels at neighboring points are 'swamped' by the addition of weighted values from the rest of the region, leaving only the major discontinuities that reflect the existence of physical boundaries in a region of space within the visual field.

There are, of course, dangers in doing this kind of smoothing-out. If one levels differences over too large a portion of the visual field, crucial details, corresponding to discernable components of what we see, will be lost. The trick then is to carry out the averaging procedure just sketched at multiple scales, and the mathematical form of the Gaussian function makes it straightforward to do this. A second issue that any realistic model of visual cognition must address is the fact that the actual mathematical operation that matches this weighted average operation holds over all points in the 'space' of the visual field - a space in which any two points may be arbitrarily close to each other. This means that no matter how many neighboring points are included in the average, there are, between any two that were included, an infinite number that have been omitted. From this point of view, simply summing the weighting of some subset of points in retinal space is inadequate. But what else is there to do?

The smoothing operation can in fact be elegantly handled by using a special mathematical technique called CONVOLUTION. Imagine that we partition one dimension of the region of space in question into a set of rectangles, one side of which has the length of the weighted intensity which we treat as constant over the width of rectangle, taken to be an extremely small fixed length written Δx . We now sum the areas of all these rectangles, allowing the width of Δx to approach arbitrarily close to zero – a standard technique in integral calculus, allowing us to calculate a number that this sum converges to no matter how small Δx gets (and therefore no matter how many rectangles of smaller and smaller width we pack together). This number is called the *limit* of the infinite sum of infinitesimally narrow rectangles, and corresponds to the weighted value of the blurred image at that point in space, with all points in the space included in the weighting. Finding this limit is carried out by taking the definite integral © JOHN BENJAMINS PUBL

of the product of the field and the weighting function over the (relevant subportion of) the visual field; again, what is critical is that infinity enters into this operation in two ways: the narrowness of each of the rectangles in the sum, and the number of rectangles which enter into that sum as their width approaches arbitrarily close to zero.

The approach to the smoothing of the visual field by what has come be called convolution with a Gaussian filter, as just described, was the first part of Marr's innovative mathematical model of early visual processing. But the second part of the edge detection task - identifying the signature properties of edges in that blurred signal - now becomes critical. As it happens, it was Marr's choice of probe for edges that allowed him to connect the mathematical model of edge detection to known properties of neuroanatomical function, making the psychophysics of vision arguably the gold standard for cognitive science and the model towards which those who characterize their research as biolinguistics aspire.

Identifying edges 2.1.3

To retrieve the discontinuities in the optical signal that the brain identifies as the physical boundaries of objects in space, Marr begins with a proposal to model edges in the via what is called the step function. Suppose we disregard the visual field for a moment and focus simply on the geometric properties of steps in a staircase. In a standard Cartesian frame of reference, where the x axis corresponds to values along a horizontal dimension, the unit step function θ is defined by

$$\theta(\mathbf{x}) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases} (0)$$

Up to a certain point in space, which we take to be the 0 value of x, the value of the function (standardly written $\theta(x)$) is 0; at that point and thereafter it has the value 1 – like an infinitely long step which appears at one point on an infinitely long floor. This is, of course, an idealization (much like the frictionless surfaces and perfectly spherical object assumed in elementary physics problems), but it does provide a kind of laboratory that allows us to test various proposals for what sorts of mathematical properties sharp discontinuities display that could serve as clues to the presence of edges in the visual signal.

Some terminology and notation are useful at this point. The rate at which a function of some variable changes its value over a vanishingly small change in the value of that variable corresponds to the operation in differential calculus of finding its first derivative. The rate of change of the position of a moving object at a single instant in time is a first derivative, called velocity; the rate of change in the value of a curve in space at a given point is called the *slope* of that curve. © JOHN BENJAMINS PUBL

A convenient way to notate the rate of change in some function f as some variable x changes is $D_x f$. We obtain this derivative straightforwardly by allowing x to change by a very small amount Δx and then comparing the ratio of the change in f to the change Δx as Δx approaches arbitrarily close to 0. In fact, $D_x f$ is nothing other than the limit of this ratio, in exactly the same sense discussed above in connection with the convolution smoothing operation.

The first striking thing about steps is that their slope – $D_x \theta$, using the notion above, which is standardly written $\delta(x)$ – is zero everywhere except at x = 0, where it is infinite. This 'spike' shape is a good model of an isolated impulse, but more useful in the present context is the behavior of this spike itself. It can be shown rigorously, but may also be apparent intuitively, that the slope of the spike – the reapplication of D_x to $\delta(x)$, which would be notated $D_x\delta(x)$, yielding the *second* derivative of the step (which can also be written $D_{rr}^2 \theta$) – has to become positively infinite, just like the spike itself, at zero. But unlike the step function itself, the slope then has to become negatively infinite as it 'falls down the other side' of the spike, like a seesaw tipping virtually instantaneously from one maximum position to the opposite one. This 180° shift in direction at the edge itself means that the line corresponding to the slope of the spike has zero amplitude at the very point where, as we have been assuming, the rise of the step itelf appears.

It follows, then, that if we take the step to correspond in the visual signal to the physical presence of an edge, then wherever the second derivative of the signal yields a zero value, we can posit the existence of an edge. Combining this method of identifying the zero crossings of the visual field with the method already described of smoothing that field out before the D_{x}^{2} operator applies to it were the two fundamental contributions Marr and his associates made to the psychophysics of vision.

But what gave Marr's model its special cachet in the history of cognitive science was the fact that the zero-crossing model has a close homologue in the specific neuroelectrical wiring of the early-to-middle visual system. This insight underwrites the special authority that Marr's work has as a touchstone for past and current 'biolinguists', who take his work to be a paradigm example of what they hope to achieve in a different sphere of cognition. Yet those who seek to reproduce Marr's success in the novel domain of natural language need to bear in mind that, by the very nature of his achievement, Marr had adopted a strongly anti-literalist ontology in his understanding of the relationship between the formal theory of visual cognition on the one hand and the neurological embodiment of that theory on the other. To see this crucial point clearly, we need to back up a bit and revisit the components of Marr's mathematical model of edge detection vis-à-vis the way © JOHN BENJAMINS PUBLISHING

From mathematical model to neural architecture 2.1.4

The first point to note is that Marr's model is based completely on methods from the domain of mathematical analysis which manipulate infinities. The Gaussian smoothing filter described earlier is based, as noted, on use of convolution integrals, where integration requires, in essence, determining the limit on the area of an infinite sum of rectangles, one side of each of which is taken to become infinitesimally small. And the D_{xx}^2 operator applied to this convolution integral involves two applications of the differential calculus technique of taking a derivative - the ratio with infinitesimally small denominator described earlier. Since it would make no sense whatever to suppose that there are cell arrays in a finite visual neuroanatomy which are iconic with the operation of taking the limit of an expression in which one variable approaches arbitrarily close to zero, or with an actual infinite sum, the obvious conclusion is that the mathematical model of edge detection Marr proposes cannot be directly instantiated in the wetware of the visual cortex.

Rather than looking for a literal neural instantiation of the mathematical model they had developed, Marr and his associates realized that to link their mathematical account with the biology of vision, they needed to identify the work that the separate components of that account were doing and identify neural complexes that would accomplish the same work. So far as the first is concerned, the key is that taking the second derivative of a Gaussian-blurred visual field turns out to be identical to taking the second derivative of the Gaussian operator, and then using that derivative to blur the raw input array. What kind of operator do we get when we take the second derivative of the Gaussian? The result of this operation is something like the original bell curve, but noticeably narrower and steeper, with negative dips on both sides that rise and flatten out to the 0 line. Basically, then, the second Gaussian derivative exaggerates the amplification of a pixel and those closest to it, but-unlike the original Gaussian-adds an inhibitory contribution to the signal via the 'dip' surrounding the central region which tends to suppress the signal from the latter. The result is that local noise over a given part of the raw signal will be levelled out, but a significant difference between the center (the excitatory portion) and the surround (the inhibitory portion) will translate into a dramatic spike in values, and a corresponding dropoff and depression of pixel values within a relatively short distance. At an actual edge, the pixels on one side of the edge will correspond to a line of positive spikes, while immediately on the other side of the edge activation line there will be a parallel 'trough', expressed in negative numbers, that appears (because of a sign reversal built into the mathematics of the convolution).

This scenario might seem problematic for a simple anatomical translation of the Marr-Hildreth scenario, however. A neuron can fire or not fire, so how can it © JOHN BENJAMINS PUBL

fire in a way that corresponds to negative numbers? What Marr realized is that the effect of the negative portion of the zero-crossing profile can be simulated by neurons firing strongly in some way which is the opposite of their immediate neighbors on the other side of the edge. Since neurons fire on an all-or-nothing basis, there can be no literal translation of the negative portion of the zero-crossing intensity profile. Something other than individual cell response seems called for.

As it happens, the physiology of the retina reflects an elegant solution to this problem. The outermost layer of retinal cells are the ordinary photoreceptors largely comprising rods and cones, but these cells are linked in clusters to a deeper layer of tissue containing what are called ganglion cells, themselves activated under one of two conditions. One possibility is that light falling in the center of a cluster of photoreceptors tends to trigger an electrical impulse in the linked ganglion cell, while light falling on the cell ring surrounding the center serves to inhibit that impulse; these ganglion cells, labeled 'on-center/off-surround', share the ganglion layer with others whose firing pattern is exactly the opposite, appropriately labeled 'off-center/on-surround'. These patterns might seem familiar: both kinds of ganglion cells correspond to the shape of Marr and Hildreth's $D_{rr}^2 G$ blurring operator (narrow, high-amplitude effect surrounded by a shallower dip in the opposite direction). The two different center/surround firing patterns are a perfect analogue of positive and negative numbers: if two of them with opposite polarity were somehow wired together and fed the same illumination, their linked actions jointly identify a shift in the direction of the amplitude.

Suppose, Marr reasoned, we measure a line of positive center/surround cells lighting up immediately across from a matching line of equally active negative cells. Such a configuration corresponds perfectly, not to the mathematical form of the $D^2_{xx}\mathcal{G}$ smoothing operator, but to its functional effect: if the two lines of cells are linked, and high activity levels on both the on-center and off-center feed higher-level neuronal structures, then we have an effective simulation of the characteristic zero-crossing property which is the signature of the mathematics of convolution under a second-order derivative operator. In Marr and Hildreth (1980), the authors suggest that such paired arrays actually do exist in the lateral geniculate nucleus, a major downstream processing structure mediating between the retina and the visual cortex.

2.2 The Marr hierarchy and neurological literalism

There are a couple of striking takeaways from this brief review of Marr's approach to the cognitive problem of visual image formation. The more obvious one is the context it provides for Marr's own expression of his attitude towards views such as © JOHN BENJAMINS PUBLISHING

To say that early visual representations are retinocentric does not literally imply that a Cartesian coordinate system, marked out in minutes of arc, is somehow laid out across the striate cortex, and that whenever some line or edge is noticed it is somehow associated with its particular x- and y- coordinates, whose values are somehow precisely carried around by the neural machinery. This process would be one way of making the representations, to be sure, but no one would seriously propose it for human vision. (Marr 1982: 42, emphasis added)

This is as explicit and emphatic a rejection of Anderson and Lightfoot's neurological literalism as one can imagine. It is probably evident, but perhaps worth stressing, that if indeed there is no explicit literal representation of a Cartesian coordinate grid in neural tissue, then there likewise cannot be any literal embodiment of mathematical structures and operations stated on those coordinates, such as differential operators or location-dependent Gaussian functions, in cortical sulci. Rather, the mathematics identifies what are, in effect, the hoops that must be jumped through by any organism or machine identifying features of the physical world by information carried in wave trains of reflected light. Precisely how those hoops are jumped through for any particular species or device is an open empirical question, but for Marr the crucial work, in the case of our species at least, and the order in which it was to be pursued, comprised the following:

- The physical problem: identify extended lines of discontinuity in space corresponding to edges and borders.
- The mathematical task: determine the step functions latent in the visual field via the zero-crossings of the Laplacian/Gaussian-convolved raw intensity array.
- The algorithm design: evaluate the zero-crossings via neighboring high (absolute-)valued difference-of-Gaussians opposed for positive and negative values respectively.³
- The wetware components: simulate the difference-of-Gaussians approximation with paired rows of on-center/off-surround and off-center/on-surround ganglion and geniculate cells.

^{3.} The difference of two Gaussians provides a robust approximation of the smoothing transformation with a second order differential operator outlined above. No derivative need actually be taken; rather, the raw image is smoothed by two simple Gaussian 'bell-curve' operators of different sizes, and then one of the two 'blurred' images is subtracted from the other. This algorithm can be done very quickly and, if the sizes chosen are correct, the result is an extremely close simulation of the action of the $D^2_{xx}\mathcal{G}$ operator on the image array. © JOHN BENJAMINS PUBL

One might reply to all this, yes, well, whatever, let's just get on with it. If biocognition - what Dana Ballard has called natural computation - is a matter of completing the analogues of Marr's four steps in whatever domain and whatever species we're interested in, fine, we won't waste time expecting theoretical representations of whatever the top-level task consists of to be directly inscribed in cytoplasm and tissue complexes and so on. We'll just go ahead and look for evidence of how the brain carries out simulations comparable in the domain of grammar to what Marr did with edge detection. This would be a considerable improvement over the Anderson-Lightfoot view of things, but it would neglect a deeper point than just the fact that probably our best-understood domain of explict cognitive modeling is totally incompatible with the literalism they advocate. The great likelihood is that the neural mechanisms which simulate the formal components of the correct theory of grammatical representation - whatever these are - will turn out to be far more difficult to identify than those involving vision. My basis for this somewhat pessimistic view is the difference between (i) the relationship between the computational problem to be solved and the organization of the neural processing stream in Marr's theory of visual perception, on the one hand, and (ii) that same relationship in the domain of natural language, on the other.

Marr's great insight was in the first place the recognition that the visual problem of identifying edges in the world involved a correspondences between spatial and neurological amplitudes, and in the second the identification of a candidate for the former which could receive a natural expression in the structure and function of particular neural complexes. Neurons specialize in expressing amplitudes; it is, in effect, what they do, via the action potentials they can propagate and the number of cells participating in that activity. There is thus a very direct relationship possible between the amplitude of signals in the visual array carried by reflected light, and the amplitude of neural activity linked (perhaps via intermediate steps) to the action of (combinations of) foveal receptors in the retina. The trick is to identify correctly just which spatial amplitudes need to be calculated and which neural mechanisms have a relation of homology to that amplitude which is predictively successful vis-à-vis the observed quantitative behavior of the visual system.

But matters are entirely different when it comes to natural languages. Unlike edges and other varieties of spatial discontinuity there is no external object at all which determine the form of the computations native to the cognitive system in question. Whereas such discontinuities are universally accepted amongst vision research scientists as a core component of how the 'sensorium of sight' operates, a consenus on what it is that the supposed analogous system for language comprises awaits us only in the far future, if ever. Grammars have been variously assumed to be conditions on phrase-structure configurations, with or without structureto-structure mapping relations; relational-arc networks; proof-theoretic calculi © JOHN BENJAMINS PUBL

defined on syntactic types with prosodic and semantic term labels; dependency graphs, and any number of other formal objects, with significant ontological differences amongst them. The mathematical interpretation of phrase markers, as discussed earlier, takes them to be partially ordered sets defined under relations of precedence and inclusion; the proofs of categorial grammar, in contrast, are inferences typically made in a some variant of a calculus of types parallel in critical respects to the implicational fragment of some substructural logic, etc. It is fair to say that even the most basic questions that a 'biolinguistics' aspiring to the scientific status of cognitive theories of vision needs to answer are, at the moment, essentially unanswered, and likely to remain so for generations.

Even more problematic, however, is the fact that the relationship between the kinds of formal objects presupposed in the framework in which virtually all self-described 'biolinguistic' research is carried out, on the one hand, and the capabilities of neurons and neural complexes on the other, is altogether obscure. The situation could not be further from the theory of visual perception, where, over the course of the past half century, different approaches have lived and died on the basis of whether or not a predictively successful connection could be established between characteristics of the object on the one hand and the functionality of the complete visual cortex on the other. Marr's computational approach, based on largely bottom-up edge detection as per the preceding discussion, played a major role in displacing the prior state-of-the-art framework, based on Fourier decomposition in separate channels, precisely because, as Westheimer notes,

> The most direct physical implementation of a spatial-frequency analysis would be by way of Fourier combs, i.e. a series of detecting templates that have a spatially sinusoidal acceptance function and whose output is proportional to the image's content at their spatial frequency. For completeness they would have to be duplicated with 90° phase shift. The eye's optics can certainly be treated in such a manner, but the anatomy of the retina is not in accord. Processing here is strictly local, transduction is compartmentalized in the structural elements, rods and cones, and the generation of neural signals is confined within small neighborhoods. (Westheimer 2001: 538; emphasis added)

He explains that, by the end of the 1970s, "the realization set in that [...] the distance between [Fourier spectral] theory on the one side, and the structure and wider actuality of visual functioning on the other, had been widening rather than narrowing" (Westheimer 2001: 538). Note also the empirical problem with Fourier spectra in terms of texture identification described in Julesz and Caelli (1979). Edge detection via differential operators applied to step functions superceded the Fourier decomposition approach precisely because a significant correlation could be made between neural function on the physiological end and mathematical © JOHN BENJAMINS PUBL

properties of the signal on the physical end, via the relationship between amplitudes noted above. In contrast, it is not in the least clear what neural function could correspond to the syntactic hierarchical structural relations and derivational changes in those structural representations assumed by essentially the entire 'biolinguistic' research community – and so far as I am aware, no concrete proposals for a plausible formal operation (corresponding to the Marr-Hildreth convolution of $D^2_{yy}G$ with the intensity array in the retina's visual field) directly implicating some neuroanatomical function (e.g. the role of retinal ganglion and lateral geniculate X-cells mirroring the firing patterns of antagonistic blocks of foveal photoreceptors) has ever been proposed in the 'biolinguistic' literature. Without such a correspondence, there cannot possibly be anything like a biology of linguistic cognition remotely comparable in any sense to the genuine biology of visual cognition that Marr's research program inaugurated, and it is therefore unremarkable, as Postal has observed, that "in total contrast to actual biological science, in four decades [Chomsky] has not specified a single physical property of any linguistic object". (Postal 2009: 256).

The bottom line, then, so far as the neurological literalism advocated by Anderson and Lightfoot is concerned, is that there is not even a glimmer on the horizon of anything suggestive of neural structures whose function corresponds to syntactic hierarchic representations, let alone relevant constraints such as c-command (or indeed any other command relation), filler/gap connectivity via multiple A movement, Late Merge or anything else in the P&P conceptual toolkit. This is the picture which emerges from the work of scientists who actually do work on the neurobiological basis of linguistic ability, as summarized in, e.g., Embick and Poeppel (2015: 358-359):

> [A]lthough cognitive theories and neurobiological theories [of natural language] are advancing in their own terms, there are few (if any) substantive linking hypotheses connecting these domains [...] There are two major problems that can be identified when we try to link C[omputational-]R[epresentational] and N[euro] B[iological] theories [...] The first is that CR and NB theories have different types of primitives, i.e., distinct ontologies, making any attempts at directly linking the two domains prima facie problematic, if not outright incoherent [...] The problem is that one cannot simply 'draw lines' between the categories provided by each domain and expect such an attempt at 'alignment' to withstand any serious scrutiny. For example, the claim that the object MORPHEME in the CR-theory corresponds to the object NEURON in the NB theory is a non-starter; it is not even wrong.

This summary speaks, I think, as directly as one could wish to the plausibility of Anderson and Lightfoot's expectation that one day an improvement in neural scanning technology will reveal the presences of (presumably binary-branching) trees somewhere in the visual cortex. The uncomfortable truth, rather, seems to © JOHN BENJAMINS PUBL

be that, as Embick and Poeppel put it, "there is at present no clear idea of how the brain represents and computes any of the computations that are part of language" (Embick & Poeppel 2015: 360).

The picture drawn by linguistic neuroscientists thus contrasts markedly with the detailed specification of comparable conceptual basics in vision research in the Marr paradigm. In this sense, 'biolinguistics' appears to be a field conspicuous mostly for having, not only no actual results to speak of, but nothing remotely close to well-defined, testable hypotheses about what kind of biological structures and mechanism to look for as embodiments of the grammatical structures its practitioners for the most part assume. If the term has any utility at all, it presumably exists, as I suggested at the outset, as an aspiration – a hope that one day a detailed connection between linguistic cognition and the detailed cortical anatomy of our species will be established that it makes sense to speak of in the same breath as the state of knowledge achieved in the realm of visual cognition forty years ago.

Domain specificity and natural language 3.

While neurological literalism might seem eccentric or bizarre even to committed adherents to one or another research program identified by its practitioners as biolinguistics (cf. Ramchand's comment and Martins and Boeckx's interesting article cited earlier), it is probably fair to say that the doctrine of domain specificity is far less likely to raise eyebrows - despite the fact that, as pointed out in Everett's excellent critical review of Anderson and Lightfoot (2002) (Everett 2005) and, still more forcefully in his evisceration of their 2006 reply in the same issue (Anderson & Lightfoot 2006, Everett 2006), there is no empirical evidence which supports this specificity over various competing alternatives. Below I sketch some research by people who are actual practicing neurocognitive scientists, published in premier journals of cortical neuroanatomy and its cognitive function, that points in quite the opposite direction. But in the following section, I present an argument that, on 'biolinguistic' assumptions - which, echoing Soames' (1984) framing of the confusion, conflate the mathematical characterization of the object of inquiry with the physical realization of tacit knowledge of that object - this specificity does *not* hold in linguistic cognition.

What computation do sets of NL sentences represent? 3.1

In addressing the general question of domain specificity, it is instructive to start from Marr's 'top level' question: what computation(s) does a theory of natural language structure require? Natural language sentences constitute pairings of © JOHN BENJAMINS PUBL

form and meanings; entailing that this pairing must be the yield of the correct theory of NL. The predominant formulations of this theory have posited recursively defined combinatoric units largely based on the narrow criteria of extraction (only constituents can undergo unbounded displacement) and proform replacement (only constituents can serve as antecedents to pro-forms). But there is strong evidence from a wide range of linguistic phenomena that there are combinatory units which must be recognized in order to achieve a credible syntax/ semantics interface, but which are neither extractable nor replaceable. On that basis, a strong case can be made that what is essentially the default assumption about syntactic representations - hierarchical organization of words into successively larger combinatory units whose structural boundaries are preserved in those representations - is not the optimal theory of syntactic structure. A few examples will illustrate the empirical basis for this scepticism about phrase markers as models of sentences.

Consider the following examples:

- (1)Chris nominated, and Terry voted for, Robin and Leslie (respectively). ≠ Chris nominated Robin and Leslie (respectively), and Terry voted for Robin and Leslie (respectively).
- (2)Robin was singing, and Leslie was whistling, the same tune. \neq Robin was singing the same tune, and Leslie was whistling the same tune.
- Robin gave the Sierra Club, and his sister pledged to the NRDC, a total of (3) \$100,000.00. \neq Robin gave the Sierra Club a total of \$100,000.00, and his sister pledged to the NRDC a total of \$100,000.00.

Such data, exemplifying what has been called Right Node Raising, exhibit systematic scope anomalies: the semantic operators corresponding to respectively, same/different/equal/identical etc., and summative predicates such as an average/total, do not distribute over their associated coordinations, but rather scope over something like the denotations of the conjoined sequences. The problem of course is that the conjoined sequences themselves are not phrase structure constituents, and hence should not be coordinable in the first place. A variety of complex operations has been proposed in a variety of phrase-structure-based approaches, all of which involve highly problematic assumptions and intricate stipulations whose objective is to somehow or other represent the coordinated material in these examples as a genuine PS constituent whose 'visible' form has been truncated by one or another means (for detailed discussion, see Levine 2011, Kubota & Levine 2014 and Kubota & Levine 2015). But in not a single case does the pattern in question fall out as the null hypothesis of *any* phrase-structure-© JOHN BENJAMINS PUBLISHING

Gapping and ellipsis phenomena exhibit the same intractability so far as the syntax/semantic interface is concerned. We have, for example, the following kinds of data (where the material missing in the second clause is notated in boldface in the first clause, and where small caps indicate contrastive emphasis):

- (4)John **bought** a CHESS set, and Mary Ø a BOOK. a.
 - I gave Mary a book, and Sue, Ø a CHESS set. [ambiguous]. b.
 - с. John decided to go with WINE, and Leslie, Ø BEER.
 - d. John continued to try to subscribe to a NEWSmagazine, and Mary, Ø to a POLITICAL journal.
- (5) You can't take the lining out of that coat. You can Ø this one. a.
 - b. I didn't try to subscribe to as many wine clubs as I did POLITICAL journals!
 - I would bet a friend more DOLLARS that the sun would rise in the WEST c. tomorrow than I would EUROS that the Earth was FLAT.

In all of these cases, what is missing, and needs to have its interpretation supplied to the remnants in the second conjunct (4) or clause (5), is not a constituent, and what actually appears in these contexts respectively does not, itself, correspond to an acceptable utterance on its own. Again, current PS-based accounts prove to be analytically unsatisfactory and empirically inadequate (see Kubota & Levine 2016a and Kubota & Levine 2017a for details). The failures of PS-based approaches across this rather vast expanse of phenomena strongly suggest that the criteria adopted in such approaches for the identification of combinatorial units are flawed at the foundations. A different kind of formal basis for syntactic representation is needed.

In general terms, the most obvious such basis is some version of Categorial Grammar (CG). As we show directly, in a particular type-logical form of CG, sequences such as gave Mary, try to subscribe to and bet a friend are all derivable as structural units, are automatically assigned a category, and at the same time are provided with both prosodic and semantic labels (which can be thought of as interpretations) leading to full compositionality in both domains.

In the past, there was much more of a consensus in theoretical linguistics that phrase structure configurations were the indispensible basis of syntactic representations. Islandhood phenomena, first extensively studied in Ross (1967), were regarded as explicable solely on the basis of the geometry of such representations, and deriving these effects from ever more abstract characterizations of syntactic structures, which arguably reached its peak in the Barriers era following the appearance of Chomsky (1986), was probably the main theoretical initiative in syntactic research. But more recent work (e.g., Kluender 1998, Hofmeister & Sag 2010, Hofmeister & Sag 2012, Hofmeister & Sag 2013a, Hofmeister & Sag 2013b, Kehler 2002, Kandybowicz 2006, Kandybowicz 2009, © JOHN BENJAMINS PUBL

Kubota & Lee 2015) has cast very strong doubt on structural sources for patterns of islandhood; see Newmeyer (2016) for a useful overview of the issues. Similarly, anaphora and related phenomena that were included under what in earlier versions of the Principles and Parameters framework was called the 'Binding Theory' were assumed to require reference to syntactic domains defined configurationally, as in Chomsky (1981), but more recent work, both within the P&P framework and outside it, no longer takes configurational relations in phrase structure trees to be critical to the definition of anaphoric possibilities (see, e.g., Pollard & Sag 1992, Pollard & Sag 1994, Jacobson 2007, Safir 2004, Nediger 2015). Moreover, the assumption of phrase structure creates, as in the case of the Gapping data presented above, and many other syntactic phenomena, a number of seemingly intractable difficulties for the syntax-semantics interface that have led to various attempts among phrase structure researchers themselves, to relax the requirement that syntactic combinatorics make reference to strictly phrase structural objects; see, e.g., McCawley (1982), Citko (2000), de Vries (2009), etc. As the apparently overwhelming evidence for configurational structure has faded and the difficulties created by the assumption of such structure as a condition on syntactic representation have become more evident, alternative approaches have become increasingly attractive. In the following discussion, based on the analysis detailed in Kubota and Levine (2016a), I lay out the basic principles of one such alternative, show how it can straightforwardly handle the syntactic/semantic difficulties posed by noncanonical varieties of coordination, and draw some conclusions based on the formalism of this approach for the notion of domain specificity in linguistic cognition.

Rules in Hybrid Type-Logical Categorial Grammar (HTLCG) 3.2

HTLCG belongs to a family of categorial grammars which can be characterized as Curryesque type logics, deductive systems isomorphic to the implicational fragment of intuitionistic linear logic. 'Curryesque' frameworks are so described on the basis of work by the logician Haskell Curry, who argued that prosodic and structural information in linguistic expressions - what he distinguished as PHENOGRAMMAR and TECTOGRAMMAR respectively - though linked, need to be kept separate and each equipped with its own proper combinatorics. This key idea was implemented in Oehrle (1994), and is reflected in the tripartite signs which appear in the proofs below; thus, the sign for give would be GIVE; give; ((NP\VP)/ NP)/NP, where GIVE is an abbreviation for a complete phonological specification, give a representation of the corresponding model-theoretic function typed $\langle e, \langle e, e \rangle$ $\langle e, t \rangle \rangle$, and the syntactic type specifies the combinatoric possibilities available to the signs bearing it and the direction in which those possibilities are realized, in accordance with the inference rules given in (6): © JOHN BENJAMI

(6)	Connective	Introduction	Elimination
	/	$\begin{split} & [\varphi; x; A]^n \\ & \vdots & \vdots \\ & \underline{b \circ \varphi; F; B} \\ & \underline{b; \lambda x.F; B/A} / I^n \end{split}$	$\frac{a; F; A/B b; G; B}{a \circ b; F(G); A} / E$
	١	$\begin{split} & [\varphi; x; A]^n \\ & \vdots & \vdots \\ & \underline{\phi \circ b; F; B} \\ & \underline{b; \lambda x.F; A \setminus B} \\ \end{split} $	$\frac{b; G; B a; F; B \setminus A}{b \circ a; F(G); A} \setminus E$
	I	$[\varphi; x; A]^{n}$ $\vdots \vdots$ $b; F; B$ $\overline{\lambda \varphi[b]; \lambda x.F; B^{\uparrow}A} I^{n}$	$\frac{a; F; A \upharpoonright B b; G; B}{a(b); F(G); A} E$

(where the \bigcirc operator denotes linear concatenation). The elimination rules are nothing other than the type-logical homologues of modus ponens in intuitionistic propositional logic. The two directional slashes are type constructors which define a category, i.e., a class of terms, which will yield a term of the type on the 'numerator' side of the slash when combined with a term matching the 'denominator' side, on that side: a *B*/*A* term concatenated with a following *A* term constitutes a *B* term, and similarly for an $A \setminus B$ term and a preceding A term. The $B \upharpoonright A$ type works the same way, except that the connective \uparrow is not directional. In effect, a $B \uparrow A$ term lacks an A term somewhere within it, and has a functional prosody and semantics which, applied respectively to the prosody and semantics of such a term, returns a sign of type B, whose prosody is derived by applying the functional prosody of $B \upharpoonright A$ to the prosody of the *B* term, and the semantics of $B \nmid A$ to the semantics of the *B* term. All three rules are type-logical correspondents of the Implication Elimination rule found in the Natural Deduction construction of standard logic in Prawitz format.

The dual of Implication Elimination, Implication Introduction, is likewise realized in the inference system in (6). The three Introduction rules given there are slightly less obvious intuitively, but in all cases the basic idea is that if by hypothesizing a type A (corresponding to a prosodic variable φ and a semantic variable x) we are able to derive a sign of type B, then by withdrawing that hypothesis we are left with a sign which *would* be of type B if it were to combine with a sign of type A. That is, if some prosodic material corresponding to φ were to appear where φ appears, and some semantic expression of the appropriate type were to appear in B's semantic term where x appears, then we would have the prosodic and semantic expressions of a type B sign. In other words, we would © JOHN BENJAMI

have a function from the prosody and semantics of an A term to the prosody and semantics of a B terms. This can be exactly modeled by taking the result of the withdrawn hypothesis to be λ -expressions abstracting on φ in the prosodic sector and on *x* in the semantic sector.

In broad terms, the rule system in (6) is a synthesis of the strictly directional syntactic calculus given in Lambek (1958) with the exclusively non-directional system developed in slightly different formulations in de Groote (2001) and Muskens (2003). The two calculi, previously regarded as mutually exclusive, jointly yield a wide variety of empirically robust results involving the interaction of, inter alia, coordination, ellipsis, and comparative constructions with scopal phenomena including generalized quantifiers, 'respectively' interpretations, symmetrical and summative predicates and much else (for details, see Kubota 2015, Kubota & Levine 2016a, Kubota & Levine 2016b, Kubota & Levine 2017a and Kubota & Levine 2017b).⁴ The interplay between directional and nondirectional inference rules allows us to saturate valence requirements, including medial arguments, with variables, and subsequently withdraw the hypotheses corresponding to these variables, allowing us to control the relationship between form and meaning over all positions in the word string. The phenomenon of Gapping provides a useful showcase for the descriptive power of this purely logic-based approach to the syntax-semantic interface.

Gapping 3.3

Gapping is a particularly mysterious instance of apparent nonconstituent coordination in which even the symmetry between the coordinated fragments apparent in Right Node Raising and Dependent Cluster Coordination is missing:

- (7)Robin speaks French, and Leslie, Ø German. a.
 - b. Robin wants to speak French, and Leslie, Ø German.
 - To Robin Chris gave the book, and to Leslie, Ø the magazine. c.

We have in each case an ostensible conjunction of a full clause with a series of phrases that appear to be remnants of a sentence from which material (possibly a nonconstituent itself) corresponding to a semantic predicate has been omitted, and whose semantics is applied to the remnant phrases in a way parallel to its application in the full conjunct on the left. There are various ways, many of them conspicuously stipulative, in which the facts in (7) can be accounted for, but any general account of Gapping also needs to be able to account for the unusual scope anomaly observed in Siegel (1984) and Oehrle (1987), exhibited in (8):

The formal soundness of HTLCG has been confirmed in Moot's demonstration that its pof theory can be embedded in first order linear logic carbon by a start of the proof theory can be embedded in first order linear logic; see Moot (to appear). © JOHN BENJAMINS PUBL

- (8)Mrs. J can't live in LA and Mr. J Ø in Boston. $(=\neg \langle [\phi \land \psi])$ a.
 - Kim **didn't play** bingo or Sandy Ø sit at home all evening. $(=\neg \langle [\phi \land \psi])$ b.

The problem is that the scope of the modal operators attached to the semantics of certain auxiliaries ordinarily does not reach beyond the confines of the minimal clause in which the auxiliaries in question appear. There have been attempts to reconcile this fact with the interpretations in question, e.g. Johnson (2000), Johnson (2009), Toosarvandani (2013), but these encounter major empirical challenges that strongly suggest the analyses offered are on the wrong track (see Kubota & Levine 2014, Kubota & Levine 2015 for detailed critiques of a range of approaches based on phrase-structural syntactic architectures, both derivational and nonderivational). The key to the solution is the recognition, following Oehrle (1987), that Gapping is in fact an instance of like-category coordination in which a single functor, corresponding in the case of (7a) to Robin_French and Mary __ German, is constructible as the conjunction of two clauses with a missing verb whose prosody is β -converted into only the first conjunct, but whose semantics is distributed over both. The kind of hypothetical reasoning used to obtain data such as (7) turns out to yield the (8) examples with no extra machinery at all.

The proof for standard Gapping is laid out in stages below. We apply the Lambek component of the system to derive a string of type S, based on a variable of type VP/NP, i.e., a normal transitive verb. The first step saturates both of the NP arguments in this variable's type description.

(9)	[φ ₁ ;	$P; VP/NP]^1$	mary; m ; N	NP /F
	john; j ; NP	$\phi_1 \circ mary;$	<i>P</i> (l); VP	— / L
	john $\circ \varphi_1 \circ Mary; P(\mathbf{m})(\mathbf{j}); S$			NT 1
	$\lambda \phi_1.john \circ \phi_1 \circ m$	ary; $\lambda P.P(\mathbf{m})$	j); S↑(VP/NP)	- 11

We now carry out a parallel proof with Bill as subject and Sue as object:

 $\lambda \varphi_1$.bill $\bigcirc \varphi_1 \bigcirc$ sue; $\lambda Q.Q(s)(b)$; S[†]TV (10)

Finally, we introduce into the proof the special conjunction operator specific to Gapping displayed in (11):

 $\lambda \sigma_2 \lambda \sigma_1 \lambda \phi_0 \sigma_1 (\phi_0) \circ \text{and} \circ \sigma_2 (\epsilon); \lambda W \lambda V. V \sqcap W; (S^{\dagger}TV)^{\dagger} (S^{\dagger}TV)^{\dagger} (S^{\dagger}TV)$ (11)

where I denotes the standard generalized conjunction operator introduced in Partee and Rooth (1983), where for propositions ϕ , ψ , $\phi \sqcap \psi = \phi \land \psi$ and for functors f, g, $f \sqcap g = \lambda \alpha f(\alpha) \sqcap g(\alpha)$. The operator in (11) differs from 'normal' and only in its phonological action: it builds an empty string argument ϵ into the prosodic func-© JOHN BENJAMINS PUBLISHING

(12)		1 1	$\lambda \sigma_2 \lambda \sigma_1 \lambda \varphi_0.$ $\sigma_1(\varphi_0) \circ and \circ \sigma_2(\epsilon);$ $\lambda W \lambda V.V \sqcap W;$ (STV)[(STV)[(STV)]	$ \begin{array}{l} \vdots & \vdots \\ \lambda \phi_1. bill \circ \phi_1 \circ sue; \\ \lambda P.P(\mathbf{s})(\mathbf{b}); \\ S^{\uparrow}TV \end{array} $	
		$\lambda \varphi_1.john \circ \varphi_1 \circ mary;$ $\lambda P.P(m)(j);$ SITV	$\lambda \sigma_1 \lambda \varphi_0. \sigma_1(\varphi_0) \circ \text{and } \circ \text{bill } \circ \epsilon \circ \text{sue};$ $\lambda V.V \sqcap \lambda P.P (\mathbf{s})(\mathbf{b});$ (SITV) (SITV)		
	met; meet; TV	$\lambda \varphi_0$ [john $\circ \varphi_0 \circ$ mary \circ and \circ bill $\circ \epsilon \circ$ sue]; $\lambda P.P(\mathbf{m})(\mathbf{j}) \sqcap \lambda Q.Q(\mathbf{s})(\mathbf{m})$; SfTV			
		john \circ met \circ mary \circ and \circ bill $\circ \boldsymbol{\epsilon} \circ$ sue;			

 $meet(m)(j) \land meet(s)(b); S$

This account extends with no further technical additions to the other examples in (7), and more complex cases still, e.g.,

(13) John solved a certain difficult crossword puzzle on Monday and Mary on Tuesday, but I don't remember what paper it was in.

Such cases fall out straightforwardly from the Gapping analysis given and the treatment of generalized quantifiers in HTLCG. But there are still more complex scoping phenomena that need to be acccounted for, e.g., those in (14):

- (14) a. John can't eat steak and Mary (just) eat pizza! $\neg \Diamond > \Lambda$
 - b. Mrs. J can't live in LA and Mr. J in Boston. $\neg \Diamond > \Lambda$
 - c. Sue wouldn't play bingo or Bill sit at home all evening. $\neg > V$

In a nutshell, such examples are instances of a higher-order version of auxiliaries (for which ample independent evidence exists) being gapped, so that the auxiliary is outside the coordinate structure, but prosodically 'lowers'. Modal wide scope then follows immediately with nothing further required. To begin with, consider

(15) Someone must have seen John. $\Box > \exists$

A lower-order entry for modal must, along the line of

(16) must; $\lambda P \lambda x. \Box P(x)$; VP/VP

will not do here. True, we can obtain a reading for (15) with a wide-scoping indefinite:



The interpretation here is that there is one particular person such that it is necessary that, in all possible worlds, that person has seen John. But this is not the sole reading. What we want is the interpretation in which what is necessary is that John was seen, period, though the speaker need have no idea by whom (more technically, in every possible world, someone saw John, but that individual isn't necessarily the same one in any two or more possible worlds). To achieve the latter reading, in which the modal necessity operator outscopes the existential, we need the higher order version of *must* given in (18):

 $\lambda \sigma.\sigma(\text{must}); \lambda \mathcal{F}.\Box \mathcal{F}(\text{id}_{\langle e,t \rangle}); S^{\dagger}(S^{\dagger}(VP/VP))$ (where $\text{id}_{\langle e,t \rangle} = \lambda P_{\langle e,t \rangle}.P$) (18)

(19)

It might seem profligate to have two separate entries for each modal, one for its narrow scope and one for its wide scope properties. A parallel duplication of quantifiers to obtain its wide-vs. narrow scope behavior, for example, is unnecessary in any formalism I know of. And in fact, such duplication is also unnecessary in the case of the modal; the HTLCG calculus derives the standard lower-order version of must as a theorem of the sign in (18) (see Kubota & Levine 2016a: 150 for a formal proof). Given this scopal operator, we can straightforwardly deduce the wide-scope interpretation of modal necessity in (15):

()			$\begin{bmatrix} \varphi_1; \\ f; \\ VP / VP \end{bmatrix}^2$	have ○ seen ○ λv. perf(see(j) VP) John; (v));	/E
		$\begin{bmatrix} \varphi_2; \\ z; \\ NP \end{bmatrix}^2$	$\phi_1 \circ hav f(\lambda v.pe VP)$	$re \circ seen \circ John rf(see(j)(v)));$	n;	-/E
			$ \varphi_1 \circ have \circ s $ $ \psi.perf(see(j)(v)) $	een ○ John;)))(z); S	$- M^2$	
	λσ.σ(someone); ∃(person); Sf(SfNP)	$\lambda \varphi_2. \varphi_2$ $\lambda z. f (\lambda$ SINP	$\circ \varphi_1 \circ have \circ v.perf(see(j)(v))$	seen ○ John;)))(z);		
	someone $\circ \varphi_1 \circ$ have \circ seen \circ John;			- 1 E		
λσ.σ(must);	$\exists (person)(\lambda z.f(\lambda v.perf(see(j)(v)))(z)); S$			\ 11		
$\lambda \mathcal{F}.\Box \mathcal{F}(\mathrm{id}_{\langle e,t\rangle});$ Sl(Sl(VP/VP))	$\lambda \varphi_1$.someone $\circ \varphi_1 \circ$ have \circ seen \circ John; $\lambda f. \exists (person)(\lambda z.f(\lambda v.perf(see(j)(v)))(z)); \$f(VP/VP)$			\F		
	someone \circ must \circ	have \circ seen	○ John;			

 $\Box \exists (person)(\lambda z.perf(see(j)(z))); S$

It turns out, however, that this independently motivated higher-order modal operator directly yields the interpretations in (14). The storyline of the proof in (20) is essentially the same as that by which interpretations of ordinary generalized quantification is obtained: a variable corresponding to a lower-order term is composed into a proof by the Lambek component of the inference rule; this variable is then © JOHN BENJAM

(20)

bound by abstraction via nondirectional implication introduction, and the resulting λ -term is provided as an argument to the higher-order operator.

(20)					
	φ ₁ ; <i>f</i> ; VP/VP	eat ○ steak; eat(s); VP			
john; j; NP		at \circ steak; ());	$\lambda \sigma_2 \lambda \sigma_1 \lambda \phi_0.$	$\vdots \vdots \\ \lambda \phi_2. mary \circ \phi_2 \circ$	
joh f(e S	$n \circ \varphi_1 \circ eat$ at(s))(j);	○ steak;	$ \begin{array}{c} \sigma_{1}(\varphi_{0}) \circ \text{and} \circ \sigma_{2}(\varepsilon); \\ \lambda \mathcal{F}_{2} \lambda \mathcal{F}_{1}.\mathcal{F}_{1} \sqcap \mathcal{F}_{2}; \\ (S^{\uparrow}X)^{\uparrow}(S^{\uparrow}X)^{\uparrow}(S^{\uparrow}X) \end{array} $	eat ○ pizza; λ <i>g.g</i> (eat(p))(m); S↑(VP/VP)	
λφ₁.jc λ <i>f.f</i> (e Sໂ(VI	$hn \circ \varphi_1 \circ eact(s))(j);$ P/VP)	at ○ steak;	$\lambda \sigma_1 \lambda \varphi_0. \sigma_1(\varphi_0)$ o and \odot $\lambda \mathcal{F}_1. \mathcal{F}_1 \sqcap \lambda g.g(eat(\mathbf{p}))($ (Sf(VP/VP))f(Sf(VP/	$mary \circ \boldsymbol{\epsilon} \circ eat \circ \mathbf{m};$ (WP))	pizza;
$\frac{\lambda \varphi_0.john \circ \varphi_0 \circ eat}{\lambda f. f(eat(s))(j) \sqcap \lambda g}$ Sf(VP/VP)		\circ steak \circ and \circ mary \circ g(eat(p))(m);	$\boldsymbol{\epsilon} \circ \operatorname{eat} \circ \operatorname{pizza};$	$\lambda \sigma. \sigma(can't);$ $\lambda \mathcal{T}. \neg \Diamond \mathcal{T}(id_{et});$ $S^{(S^{(VP/VP))}}$	
		johr ¬≬[o S	$n \circ can't \circ eat \circ steak \circ eat(s)(j) \wedge eat(p)(m)];$	and \circ mary $\circ \epsilon \circ \epsilon$	at \circ pizza;

The wide-scope interpretation of the modal in (14a) is thus accounted for, requiring only the higher-order modal and the gapping operator, both of which are needed for reasons quite independent of the Oehrle-Siegel phenomenon. And as shown in Kubota and Levine (2016a), all of the other Oehrle-Siegel anomalous scoping cases in Gapping fall out as well, on the same basis. As noted above, the HTLCG framework makes available elegant and comprehensive accounts of the syntax/semantics interface for Gapping phenomena, the interaction of scopal operators such as generalized quantifiers with 'nonconstituent' coordination (Kubota & Levine 2015, Kubota & Levine 2016a), respectively readings, symmetrical predicates and summative predicates, their interaction with each other and with 'nonconstituent' coordinations (Kubota & Levine 2016b; cf. (21a)), and pseudogapping and its interactions with, e.g., Gapping (Kubota & Levine 2017a; cf. (21b)).

- (21)a. John offered, and Mary gave, the same advice to Bill and Anne (respectively) on the same day.
 - John can eat more pizza than Bill can sushi or Mary fish & chips. b.

Given the success of HTLCG providing completely explicit accounts of this rather large range of notoriously problematic empirical phenomena on the basis of a simple formal calculus without any appeal to extra stipulations, inexplicit or even unformulated principles, or violations of basic compositionality, I think it can © JOHN BENJAMINS PUBL

fairly claim some credibility as a candidate 'optimal theory' of NL. The critical point of this whole section, however, is that this broad and deep coverage has been achieved on the basis of a combinatoric system consisting of nothing more than a type-logical homology with the implicational fragment of intuitionistic propositional logic. A careful inspection of the preceding proofs will make clear that every proof step in every proof belongs to either the set of introduction rules or the set of elimination rules for the connectives instantiating hypothetical reasoning and modus ponens respectively in the hybrid proof theory that serves as the platform for HTLCG - a fact with significant consequences for 'biolinguistic' claims of domain specificity.

Implication is not domain specific 4.

Suppose then that the combinatorics of the syntax/semantics interface in natural language takes the form of a fairly literal version of the inference rules in some standard logic. The most important takeaway from this possibility is, as I now argue, that there is no secure basis for any claim of biologically determined domain specificity in determining the human capacity for language. To establish this point, I first consider the relationship between the cognitive basis of ordinary reasoning, on the one hand, and the formal rules of logical calculi on the other.

The psychologist John Macnamara has some useful comments on this question. He observes (Macnamara 1986: 31-34) that

> The main task for the psychologist who is studying human reasoning is to account for both our ability to reason validly and our intuitions about logical validity. The set of valid inferences is infinite [...] it follows that we must have access to a set of rules that can be combined in various ways to yield an infinite set of inferences. Thus the foundations of the logics at which logicians aim, the ideal logic(s), must be psychologically real in the sense of being instantiated in some form in the mind. Further the best logics found in logic books today provide the best available guides to logical competence. It does not follow, of course, even if a logic characterizes an aspect of logical competence, that the competence is instantiated in any of its currently available forms, say, axiomatic or natural deduction.

> Basic logical competence [...] include[s] the logical resources that are deployed in natural-language sentences. These include some set of sentential connectives [...] [whose] logic [...] in its simplest aspects may be given to us in a form that is close to natural deduction rules.

(Emphases added). The point is that sets of logical inference rules in explicit deductive calculi are mathematical systems which can usefully model psychological constraints on acceptable sequences of thoughts in the course of everyday © JOHN BENJAMINS PUBL

reasoning. These systems are in effect distillations of that reasoning which capture their invariant properties over all possible content. The psychology of reasoning is an ontologically quite different creature from the rigorous proof theories which constitute that distillation, but in effect simulates the reasoning steps imposed by these proof theories – particularly in their natural deductive formulation.

Rips (1994) takes this programmatic view of the relationship between logical cognition and its mathematical formalization as his point of departure, seting forth a detailed model of reasoning, with extensive experimental support, built around a core of natural deduction rules and elementary theorems of those rules called PSYCOP (psychology of proof). At the heart of his proposal is the premise that

> a person faced with a task involving deduction attempts to carry it out through a series of steps that take him or her from an initial description of the problem to its solution. These intermediate steps are licensed by mental inference rules, such as modus ponens, whose output people find intuitively obvious.

> > (Rips 1994: x., emphasis added)

Rips argues that a psychological instantiation of natural deduction inference rules and elementary entailments from those rules undergirds deductive reasoning generally, and that "deduction must be a component of many other psychological processes on any reasonable account of their operation" (Rips 1994: 12), citing comprehension, planning and several other higher-order cognitive activities. He further observes that the reasoning process summarized in the rules for the introduction and elimination of the connectives in standard logics must be built into the neural basis of thinking, pointing out that without these fundamental reasoning steps at the outset, there is no way that learning could take place, because these reasoning steps are crucial (and prior) to any conclusion that could be drawn from experience.

Suppose then that it is some transduction of standard deductive calculi into type-logical systems in which (some subportion of) the same rules of inference are faithfully maintained which constitutes the combinatoric core of natural language capability - as we would certainly conclude if some type-logical grammar such as HTLCG (or one of the several alternative logic-based versions of Categorial Grammar) proves to be a successful competitor in the marketplace of grammatical frameworks. Suppose further that the reasoning methods embodied in those calculi are also employed across the spectrum of higher-order cognitive activities, as much of Rips' survey of experimental evidence argues for. Then the crucial premise of domain specificity for linguistic knowledge, and the idea that 'biolinguistics' has as its object a circumscribed set of neural structures dedicated to a cognitively unique system of functions, become radically untenable. © JOHN BENJAMINS PUBLI

This conclusion receives some support from recent work on the neurobiology of language carried out by actual neurobiologists - what one might call real biolinguistics, without the need for scare quotes. Particularly interesting are the results of the experimental study reported in Reverberi et al. (2007), which investigated what the authors call 'elementary inferences' - simple deductions based only on implication and disjunction elimination. Using functional magnetic resonance imaging technology to identify centers of brain activation during subjects' execution of the inference tasks presented during the experiments, they were able to demonstrate that simple deductions were associated most strongly with a specific complex of neural structures:

> Our results implicate the left inferior frontal gyrus (mainly B[rodmann's]A[rea] 44, left precentral gyrus (BA 6) and left parietal cortex (BA 40)) [...] Based on this evidence, we argue that the frontoparietal network identified represents the neural basis of the generation of conclusion in elementary deductive problems. (Reverberi et al. 2007: 758)

BA 44 is in fact none other than the celebrated Broca's area, while BA 40 is a component of what Ardila et. al (2016) identify as part of a peripheral zone in an "extended Wernicke's area", contrasting with the "Broca's complex" they identify as "a complex frontal subcortical circuit involved in language production and grammar" (Ardila et al. 2016: 120). The rather striking fact which emerges from Reverberi et al.'s results is that the core area involved in the most basic kind of deduction - modus ponens, whose dual is hypothetical reasoning - is located in what is currently regarded as the neural center of grammatical cognition. These results are highly suggestive in terms of my proposal that grammatical combinatorics are nothing other than a version of logical inference restricted to the implicational fragment of standard logics.

It might be objected that, since the nature of the deductions that Reverberi et al's experimental subjects were asked to carry were framed as verbal problems, and since BA 44 is perhaps the key area for the grammatical processing of oral language, the high level of activity measured in this region of the cortex is unsurprising. But this objection would be misguided, for, as reported in Monti et al. (2009), a set of laboratory trials revealed that verbally framed logical deductions of a relatively high degree of complexity, corresponding in difficulty to the De Morgan laws "did not recruit [regions typically reported for linguistic processing] but rather a network of regions highly similar to that reported in previous studies of deduction with sentential connectives and quantifiers" (Monti et al. 2009: 12555), emphasizing later in their paper that "inference involving sentential connectives relies on a circuit that is largely independent of areas recruited by semantic and syntactic processes specific to natural language". Monti et al. note the apparent discrepancy between their results and Reverberi et al.'s, suggesting that it is the very simplicity © JOHN BENJAMINS PUBL

of the deductive requirements in the latter's study which leads to these divergent patterns; as they note, "[t]he more challenging deductions figuring in the present experiment provoke extended and vigorous reasoning, not to be expected from elementary schemata like modus ponens" (Monti et al. 2009: 12557).⁵

Further evidence against domain specificity from real biolinguistics comes from the work of Tettamanti et al. (2009), who observed that Brodmann's area 44 is implicated in syntactic computations in a separate cognitive module, the visuospatial domain. Specifically, Tettamanti et al. found that the processing of what they call N[on]R[igid]S[equential][D]ependencies - essentially, relationships which are defined by arbitrary iterations of co-constituency definitions – are in all cases associated with Brodmann's area 44. They found that

> non-rigidly organized stimuli in both the language and the visuo-spatial domain are processed by a common bilateral fronto-parietal network, with an essential contribution of the left I[nferior]F[rontal]G[yrus]. In other words, in the presence of non-rigid dependencies, the processing of spatial information also depends on left hemispheric recruitment, in a qualitatively similar way to the processing of linguistic information [...] More specifically, it was recently shown that the left IFG is crucially involved in the executive control of hierarchically organized action sequences. (Tettamanti et al. 2009: 836.)

5. It is worth noting that once one gets very far away from the kinds of elementary reasoning that Reverberi et al. based their test suite on, the actually reasoning steps that people use to determine whether a given inference is valid almost certainly look very little like the tidy proofs that count as formally correct, notwithstanding the appellation 'natural deduction'. It is unlikely that an ordinary reasoner, confronted by the premises 'Either John was out late or Mary was home early' and 'John wasn't out late', concludes that Mary must have been home early by appeal to the rule of Disjunction Elimination and a subproof of the radically counterintuitive ex falso quodlibet conclusion that $p \vdash \neg p \supset q$, a crucial part of arriving at the immediately obvious inference. It is quite plausible, in fact, that what Monti et al. observed was the operation not of literal logical inference in the deduction of complex theorems such as the De Morgan laws, but rather the kind of ordinary reasoning in which each step involves the chunking together of several basic steps into what we might think of as macros. This aspect of the reasoning process is still largely unknown, but it would not be at all surprising to find that it does not involve the same circuitry as that simple elementary applications of the rules for implication that Reverberi et al. found linked to the Brodmann's language areas connected with grammatical knowledge, and which I have suggested above represent a serious candidate for the formal representation of the combinatorics determining the syntax-semantics interface. It seems quite possible, in fact, that ordinary reasoners do not approach the John/Mary example by some simulation of a proof of the Disjunctive Syllogism, but rather access something much more like the truth table for v, which immediately guarantees that if one disjunct is false, the disjunction can only be true if the other disjunct is. That is, the semantics of or may in effect prepackage the Disjunctive Syllogism without any deductive reasoning being involved at all. © JOHN BENJAMINS PUBL

They conclude that "the human brain has some distinctive traits by which it is capable of encoding NRSD across diverse higher cognitive functions" (Tettamanti et al. 2009: 836).⁶ Once again, then, we find that research on the biology of language which is informed by actual knowledge of biology yields results that severely challenge the 'encapsulated modularity' view of human linguistic cognition asserted in, e.g., Fodor (1983).

Summary and conclusions 5.

The preceding discussion has I think shown that there is a crushingly heavy burden of proof on advocates of neuroanatomical literalism as a premise of the 'biolinguistic' program: the infinities in the theory of limits alone make the visual system inherently uninstantiable in literal fashion in the brain, on the assumption that Marr's analysis (or something like it, e.g, a neurological expression of the Canny edge-detection model, which takes the gradient of a smooth-filtered visual array as its core computation) is correct. And there is no more reason to believe that the discrete infinities that so much has been made of in previous discussion can be any more easily be embodied in 1-to-1 fashion in the fine-grained anatomy of the cortex.

Even Chomsky himself appears to have finally accepted this conclusion, judging by his remarks a few years ago in The Science of Language:

> In the work that I've done since *The Logical Structure of Linguistic Theory* – which just assumes set theory - I would think that in a biolinguistic framework you have to explain what that means. We don't have sets in our heads. So you have to know that when we develop a theory about our thinking, about our computation, internal processing and so on in terms of sets, that its going to have to be translated into some *terms that are neurologically realizable.* (Chomsky 2012: 91, emphasis added)

- a flat contradiction to Chomsky's biologically bizarre saltation hypothesis whereby the ability to iterate the set-building operations underlying natural language capabilities was the result of a single genetic mutation in a single individual in Homo sapiens' ancestral chain. For © JOHN BENJAMINS PUBL a penetrating critique of the scientific status of these speculations, see Behme (2014).

^{6.} Tettramanti et al. also make the extremely interesting point that

the fact that [...] some nonhuman species can be taught simple NRSD is consistent with the view that language emerged in the course of evolution by drawing on set of cognitive and computational capabilities that, at least in rudimentary form, are shared across higher vertebrates. (Tettamanti et al. 2009: 837)

Apparently, then, Chomsky himself now disagrees with Anderson, Lightfoot and other neural literalists about just what we might expect to see as cortical imaging techniques continue to improve.

But even if this critical point is recognized, what remains is a formidable obstacle that does not seem to have been acknowledged even in work such as Martins and Boeckx (2016), which reveals at least some awareness of the huge gap between what has been done and what would need to be done to justify identifying biolinguistics as an actual science based on a range of robust results. The critical problem, again, is the fact in order to bridge this gap, researchers have to agree on just what it is that their best computational-level models - the abstract objects that are presumably realized in neural wetware - are models of.

The crucial contrast is with the detailed mathematical model of early visual processing discussed in § 2 above, which Marr and his associates were able to work out based on the key insight that the inital step in visual scene construction as a cognitive activity was the recognition of spatial discontinuities, and the possibility of treating the latter in terms of step functions, using the tools of signal processing technology to define a background against which these discontinuities would be recoverable using the methods outlined earlier. The empirical gravitas of that enterprise was underwritten by specific predictions about the information content of grey-level arrays subject to the convolution and zero-crossing calculations detailed in the Marr-Eldreth theory of edge detection and its neuroanatomical expression; likewise for higher order operations proposed by others in the Visionaries group of researchers at MIT in the early 1980s and after.

In contrast, the 'biolinguistics' community has yet to achieve a secure identification of any computational task in the domain of language comparable to the Marr group's edge detection hypotheses. Marr's breakthrough was in correctly identifying, at the first level of approximation, the crucial 'object of computation' for visual cognition. What is the corresponding object in the domain of linguistic cognition? Right now, we cannot talk about anything finer-grained than Brodmann's areas; to identify specific neural structures will require an insight fully comparable to Marr's forty years ago, and quite possibly still deeper. And despite all of the aspirational enthusiasm - and outright boasting - in the 'biolinguistic' literature, it seems very unlikely that anyone in that research community would claim that we are on the verge of such a breakthrough, let alone that we have already achieved anything remotely like it.

This is where the issue of domain specificity becomes crucial. While the implementation of at least certain components of visual computation does appear to be a highly specific, dedicated system unconnected with other sensory modalities, the authentically biolinguistic work cited above – carried out, it should be stressed, by actual experts in the biology of the human nervous system - has instead provided © JOHN BENJAMINS PUBL

very suggestive evidence of domain generality, with the real-time implementations of basic deductive reasoning, sentence processing and - as per the research reported in Tettamanti et al. (2009) - visual/spatial awareness, all sharing the same circuitry, in Brodmann Area 44 and other neurological domains.

In view of these severe obstacles to the prospect of identifying distinctively neurolinguistic structures in the human neocortex, it seems appropriate to conclude with some observations by a prominent proponent of the 'biolinguistic' perspective that seem applicable to where things stand, and are likely to continue to stand, into at least the foreseeable future:

> The reasons why economics is unlikely to reduce to physics are paralleled by those which suggest that psychology is unlikely to reduce to neurology. There are no firm data for anything but the grossest correspondence between types of psychological states and types of neurological states, and it is entirely possible that the nervous system of higher organisms characteristically achieves a given psychological end by a wide variety of neurological means. It is also possible that given neurological structures subserve many different psychological functions at different times [...] In either event, the attempt to pair neurological structures which psychological functions could expect only limited success. (Fodor 1975: 17, emphasis added)

All that is needed to capture the current state of 'biolinguistics' is to replace 'psychology' in the above passage with 'linguistics' – and, after all, Chomsky has for by far the better part of his career, identified linguistics as indeed a branch of 'theoretical psychology'. In the more than forty years since Fodor's comment appeared, the substantive situation with respect to 'biolinguistics' has not changed in the slightest, as Embick and Poeppel's state of the art overview cited earlier attests - notwithstanding the 'triumphalistic rhetoric' alluded to in a passage from Newmeyer (2003: 586) devoted to Chomsky's characterizations of minimalism, but also applicable to the tone of the neural literalism that Ramchand found so objectionable. And there is nothing on the horizon to lead us to expect that situation to change.

References

- Anderson, Stephen R. & Lightfoot, David. 2002. The Language Organ: Linguistics as Cognitive Physiology. Cambridge: CUP. doi:10.1017/CBO9780511613869
- Anderson, Stephen R. & Lightfoot, David. 2006. Biology and language: A response to Everett (2005). Journal of Linguistics 42: 377-383. doi:10.1017/S002222670600394X
- Ardila, Alfredo, Bernal, Byron & Rosselli, Monica. 2016. How localized are language brain areas? A review of Brodmann areas' involvement in oral language. Archives of Clinical
- Behme, Christina. 2012. The essence of Postal's criticism: A short reply to Ulfsbjorninn (2012). http://ling.auf.net/lingburg/001573 © JOHN BENJAMINS PUBLISHING

Behme, Christina. 2013. Biolinguistic platonism remains an oxymoron. http://ling.auf.net/ lingbuzz/001765>

Behme, Christina. 2014. A galilean science of language. Journal of Linguistics 51: 671-704. doi: 10.1017/S0022226714000061

Chomsky, Noam. 1976. On the nature of language. In Origins and Evolution of Language and Speech, Steven R. Harnad, Horst D. Steklis & Jane Lancaster (eds). Annals of the New York Academy of Sciences 280: 46-57. doi:10.1111/j.1749-6632.1976.tb25470.x

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

- Chomsky, Noam. 1983. Some conceptual shifts in the study of language. In How Many Questions? Essays in Honor of Sidney Morgenbesser, Leigh S. Cauman, Isaac Levi, Charles D. Parsons & Robert Schwartz (eds), 154-169. Indianapolis IN: Hackett.
- Chomsky, Noam. 1986. Barriers. Cambridge MA: The MIT Press.
- Chomsky, Noam. 2012. The Science of Language. Cambridge: CUP. doi: 10.1017/CBO9781139061018
- Citko, Barbara. 2000. Parallel Merge and the Syntax of Free Relatives. PhD dissertation, SUNY.
- de Groote, Philippe. 2001. Towards abstract categorial grammars. In Association for Computational Linguistics, 39th Annual Meeting and 10th Conference of the European Chapter, 148 - 155.
- de Vries, Mark. 2009. On multidominance and linearization. Biolinguistics 3(4):344-403.
- Embick, David & Poeppel, David. 2015. Towards a computational(ist) neurobiology of language: Correlational, integrated and explanatory neurolinguistics. Language, Cognition and Neuroscience 30: 357-366. doi:10.1080/23273798.2014.980750
- Everett, Daniel L. 2001. Discussion: on nonobjects of linguistic study. Linguist List 12(i8i6). <http://linguistlist.org/issues/12/12-I8i6.html>

Everett, Daniel. 2005. Biology and language: A consideration of alternatives. Journal of Linguistics 41:157-175. doi:10.1017/S0022226704003093

- Everett, Daniel. 2006. Biology and language: Response to Anderson & Lightfoot. Journal of Linguistics 42(2): 385-393. doi:10.1017/S0022226706003951
- Fodor, Jerry. 1975. The Language of Thought. Cambridge MA: Harvard University Press.
- Fodor, Jerry. 1983. The Modularity of Mind. Cambridge MA: The MIT Press.
- Hofmeister, Philip, Staum Casasanto, Laura & Sag, Ivan A. 2012. Misapplying working memory tests: A reductio ad absurdum. Language 88: 408-409. doi:10.1353/lan.2012.0033
- Hofmeister, Philip, Staum Casasanto, Laura & Sag, Ivan A. 2013a. Islands in the grammar? Standards of evidence. In Experimental Syntax and Island Effects, Jon Sprouse & Norbert Hornstein (eds), 42-63. Cambridge: CUP. doi:10.1017/CBO9781139035309.004
- Hofmeister, Philip, Arnon, Inbal, Jaeger, T. Florian, Sag, Ivan & Snider, Neal. 2013b. The source ambiguity problem: Distinguishing the effects of grammar and processing on acceptability judgments. Language and Cognitive Processes 28: 48-87.
- Hofmeister, Philip & Sag, Ivan A. 2010. Cognitive constraints and island effects. Language 86(2): 366-415. doi:10.1353/lan.0.0223
- Jacobson, Pauline. 2007. Direct compositionality and variable-free semantics: The case of "Principle B" effects. In Direct Compositionality, Chris Barker & Pauline Jacobson (eds.), 191-236. Oxford: OUP.
- Johnson, Kyle. 2000. Few dogs eat Whiskas or cats Alpo. In University of Massachusetts Occasional Papers, vol. 23, Kiyomi Kusumoto & Elisabeth Villalta (eds), 47-60. Amherst MA: GLSA.

Johnson, Kyle. 2009. Gapping is not (VP-) ellipsis. Linguistic Inquiry 40(2): 289-328. © JOHN BENJAMINS doi: 10.1162/ling.2009.40.2.289

- Julesz, Bela & Caelli, Terry. 1979. On the limits of Fourier decompositions in visual texture perception. Perception 8: 69-73. doi:10.1068/p080069
- Kandybowicz, Jason. 2006. Comp-trace effects explained away. In Proceedings of WCCFL 25, 220-228. Somerville MA: Cascadilla Proceedings Project.
- Kandybowicz, Jason. 2009. Embracing edges: Syntactic and phono-syntactic edge sensitivity in nupe. NLLT 27: 305-344.

Katz, Jerrold J. & Postal, Paul M. 1991. Realism vs. conceptualism in linguistics. Linguistics and Philosophy 14: 515-554. doi:10.1007/BF00632596

Katz, Jerrold J. 1981. Language and Other Abstract Objects. Totowa, NJ: Rowman and Littlefield.

Kehler, Andrew. 2002. Coherence, Reference and the Theory of Grammar. Stanford CA: CSLI.

- Kluender, Robert. 1998. On the distinction between strong and weak islands: A processing perspective. In The Limits of Syntax (Syntax and Semantics 29), Peter Culicover & Louise McNally (eds), 241-279. San Diego CA: Academic Press.
- Kubota, Yusuke. 2015. Nonconstituent coordination in Japanese as constituent coordination: An analysis in Hybrid Type-Logical Categorial Grammar. Linguistic Inquiry 46(1):1-42. doi:10.1162/LING_a_00174
- Kubota, Yusuke & Lee, Jungmee. 2015. The Coordinate Structure Constraint as a discourse oriented principle: Further evidence from Japanese and Korean. Language 91(3): 642-675.
- Kubota, Yusuke & Levine, Robert. 2014. Scope anomaly of gapping. In Proceedings of NELS 44, 247-260. Amherst MA: GLSA.
- Kubota, Yusuke & Levine, Robert. 2015. Against ellipsis: Arguments for the direct licensing of 'non-canonical' coordinations. Linguistics and Philosophy 38(6): 521-576. doi: 10.1007/s10988-015-9179-7
- Kubota, Yusuke & Levine, Robert. 2016a. Gapping as hypothetical reasoning. Natural Language and Linguistic Theory 34(1):107-156. doi:10.1007/s11049-015-9298-4.
- Kubota, Yusuke & Levine, Robert. 2016b. The syntax-semantics interface of 'respective' predication: A unified analysis in Hybrid Type-Logical Categorial Grammar. Natural Language and Linguistic Theory 34(3): 911–973.
- Kubota, Yusuke & Levine, Robert. 2017a. Pseudogapping as pseudo-VP ellipsis. Linguistic Inquiry 48(2): 213–257.
- Kubota, Yusuke & Levine, Robert. 2017b. Scope parallelism in coordination in Dependent Type Semantics. In New Frontiers in Artificial Intelligence, Mihoko Otake, Setsuya Kurahashi, Yuiko Ota, Ken Satoh & Daisuke Bekki (eds.), 79-92. Berlin: Springer.
- Lambek, Joachim. 1958. The mathematics of sentence structure. American Mathematical Monthly 65(3):154-170. doi:10.2307/2310058
- Leipnik, Roy B. 1960. The extended entropy uncertainty principle. Information and Control 3(1):18-25. doi:10.1016/S0019-9958(60)90227-8
- Levine, Robert. 2011. Linerarization and its discontents. In The Proceedings of the 18th International Conference on Head-Driven Phrase Structure Grammar, Stefan Müller (ed.), 126-146. Stanford CA: CSLI.
- Macnamara, John. 1986. A Border Dispute: the Place of Logic in Psychology. Cambridge MA: The MIT Press.
- Marr, David. 1982. Vision. Cambridge MA: The MIT Press.
- Marr, David & Hildreth, Ellen. 1980. Theory of edge detection. Proceedings of the Royal Society
- Martins, Pedro Tiago & Boeckx, Cedric. 2016. What we talk about when we talk about biolinguis-tics. *Linguistics Vanguard* 2(1). https://www.domnut.com lingvan-2016-0007/lingvan-2016-0007.xml> © JOHN BENJAM

- McCawley, James D. 1982. Parentheticals and discontinuous constituent structure. Linguistic Inquiry 13(1): 91–106.
- Monti, Martin M., Parsons, Lawrence & Osherson, Daniel N. 2009. The boundaries of language and thought in deductive inference. Proceedings of the National Academy of Sciences 106: 12554-12559. doi:10.1073/pnas.0902422106
- Moot, Richard. To appear. Hybrid type-logical grammars, first-order linear logic and the descriptive inadequacy of Lambda grammars. Journal of Logics and their Applications (International Federation of Computational Logic).
- Muskens, Reinhard. 2003. Language, lambdas, and logic. In Resource Sensitivity in Binding and Anaphora, Geert-Jan Kruijff & Richard Oehrle (eds), 23-54. Dordrecht: Kluwer. doi: 10.1007/978-94-010-0037-6 2
- Nediger, Will. 2015. Focus and Condition C. In Proceedings of the 32nd West Coast Conference on Formal Linguistics, Ulrike Steindl, Thomas Borer, Huilin Fang, Alfredo García Pardo, Peter Guekguezian, Brian Hsu, Charlie O'Hara & Iris Chuoying Ouyang (eds), 296-302. Somerville MA: Cascadilla Proceedings Project.
- Newmeyer, Frederick J. 2003. Review article on On Nature and Language by Noam Chomsky, The Language Organ: Linguistics as Cognitive Psychology by Stephen Anderson and David Lightfoot, and Language in a Darwinian Perspective by B. H. Bichakjian. Language 79: 583-599. doi:10.1353/lan.2003.0180
- Newmeyer, Frederick J. 2005. Possible and Probable Languages: A Generative Perspective on Linguistics. Oxford: OUP. doi:10.1093/acprof:0s0/9780199274338.001.0001
- Newmeyer, Frederick J. 2016. Nonsyntactic explanations of island constraints. Annual Review of Linguistics 2: 187-210. doi:10.1146/annurev-linguistics-011415-040707
- Oehrle, Richard T. 1987. Boolean properties in the analysis of gapping. In Discontinuous Constituency (Syntax and Semantics 20), Geoffrey J. Huck & Almerindo E. Ojeda (eds), 203-240. New York NY: Academic Press.
- Oehrle, Richard T. 1994. Term-labeled categorial type systems. Linguistics and Philosophy 17(6):633-678. doi:10.1007/BF00985321
- Partee, Barbara & Rooth, Mats. 1983. Generalized conjunction and type ambiguity. In Meaning, Use, and Interpretation of Language, Rainer Bäuerle, Christoph Schwarze & Arnim von Stechow (eds), 361-383. Berlin: Walter de Gruyter. doi: 10.1515/9783110852820.361
- Pollard, Carl & Sag, Ivan A. 1992. Anaphors in English and the scope of binding theory. Linguis*tic Inquiry* 23(2): 261–303.
- Pollard, Carl & Sag, Ivan A. 1994. Head-Driven Phrase Structure Grammar. Studies in Contemporary Linguistics. Chicago IL: University of Chicago Press.
- Postal, Paul M. 2009. The incoherence of Chomsky's 'biolinguistic' ontology. Biolinguistics 3(1):247-266.
- Reverberi, Carlo, Cherubini, Paolo, Rapisarda, Attilio, Rigamonti, Elisa, Caltagirone, Carlo, Frackowiak, Richard S. J., Macaluso, Emiliano & Paulesu, Eraldo. 2007. Neural basis of generation of conclusions in elementary deduction. NeuroImage 38: 752-762. doi: 10.1016/j.neuroimage.2007.07.060
- Rips, Lance J. 1994. The Psychology of Proof. Cambridge MA: The MIT Press.
- Ross, John Robert. 1967. Constraints on Variables in Syntax. PhD dissertation, MIT. Reproduced by the Indiana University Linguistics Club.
- © JOHN BENJAMINS PUBLISHING COMPANY

Siegel, Muffy E. A. 1984. Gapping and interpretation. Linguistic Inquiry 15(3): 523-530.

- Soames, Scott. 1984. Linguistics and psychology. *Linguistics and Philosophy* 7:155–179. doi:10.1007/BF00630811
- Tettamanti, Marco, Rotondi, Irene, Perani, Daniella, Scotti, Giuseppe, Fazio, Ferrucio, Cappa, Stefano F. & Moro, Andrea. 2009. Syntax without language: neurobiological evidence for cross-domain syntactic computations. *Cortex* 45:825–838. doi:10.1016/j.cortex.2008.11.014
- Toosarvandani, Maziar. 2013. Gapping is low coordination (plus VP-ellipsis): A reply to Johnson. Ms, MIT.
- Westheimer, Gerald. 2001. The Fourier theory of vision. *Perception* 30: 531–541. doi:10.1068/p3193

