biolinguistics

XP

CP

...

γ

α

XP

The Volume 4 *Biolinguistics* Editorial

Kleanthes K. Grohmann & Cedric Boeckx

The fourth volume of *Biolinguistics* signals a new road along several dimensions. First and foremost, of course, the fact that we have been able to compile these four volumes, with a total of nearly one thousand pages (a single-issue volume 1 of 150 pages and three-issue volumes 2 and 3 of 364 and 420 pages, respectively, which each had a double-issue, the latter in the form of a special issue), is a feast in and of itself. It means that we as editors have survived, that the journal is on a very good start indeed, and that, more generally, *Biolinguistics* generates enough interest in the relevant communities to sustain publication—all of which would not have been possible without the support of you, the readers, the reviewers, the members that serve on the various boards, and of course, the respective authors. So, first of all, many thanks to the entire supporting *Biolinguistics* community.

Second, as editors we can detect a slight move in the submissions from core theoretical issues in linguistics to wider cognitive and biological issues in language, that is, the kind of interdisciplinarity we aimed for originally and would like to continue strengthening. For this purpose, we decided to rotate and reshuffle our main supporting means—the Advisory Board, the Editorial Board, and the (student) Task Team—in order to reflect the increasingly new demands and thereby making ties with communities of researchers outside (theoretical) linguistics investigating the nature and structure of language.

We would thus like to use this opportunity to thank all outgoing board members for the services they have provided over the past four years, which include the dreaded but vital chore of reviewing manuscript submissions, help with the acquisition of new submissions, and general support and advertising of the journal among their respective circles as well as pointing out specific issues and suggestions to us, the editors. These colleagues are listed here in alphabetical order:

Joseph Aoun	Howard Lasnik
Hagit Borer	Jason Merchant
Željko Bošković	Jairo Nunes
Michael Brody	E. Phoevos Panagiotidis
Gennaro Chierchia	Anna Roussou
Elan B. Dresher	Joachim Sabel
Mark Hale	George Tsoulas
C.-T. James Huang	Jochen Zeller
Manfred Krifka	

At the same time, we would like to welcome the new members on the *Biolinguistics* boards. This goes for all colleagues who now serve on the Advisory Board, some of whom have already served a term on the Editorial Board:

Stephen R. Anderson	James Hurford
Anna Maria Di Sciullo	Lyle Jenkins
W. Tecumseh Fitch	Massimo Piattelli-Palmarini
Naoki Fukui	

The same warm welcome goes to the new members of the Editorial Board:

Sharon Armon-Lotem	Guillermo Lorenzo
Antonio Benítez-Burraco	Carme Picallo
Robert C. Berwick	Friedemann Pulvermüller
Ina Bornkessel-Schlesewsky	Charles Reiss
Albert Costa	Bridget Samuels
Simon E. Fisher	Tobias Scheer
Naama Friedmann	Alessandro Treves
Koji Fujita	Robert Truswell
Itziar Laka	

In addition, *Biolinguistics* is supported by a Task Team consisting of under- and postgraduate students. We thank Jeffrey K. Parrott, Bridget Samuels, and Adam Szczegielniak for having served on the Task Team who, in the meantime, received their doctorate as well as the other members who served us on the Task Team at some time or other over the past four years: Michael Beys, Marina Iakovou, Clemens Mayr, and Philip Rausch. The reshuffled Task Team has now been assigned more specific tasks, including setting up, maintaining, and feeding the official *Biolinguistics* web log. We thus welcome Evelina Leivada and Natalia Pavlou, who support the operating editor for much needed editorial assistance, such as pre-formatting accepted manuscripts for publication, and we extend Task Team tasks to Txuss Martín and Hiroki Narita with support from Terje Lohndal and Dennis Ott (under Bridget Samuels' supervision), who are now responsible for managing the *Biolinguistics* blog.

This blog is another innovation that comes with volume 4. In early 2010, it was set up for spreading information, announcing relevant events, and—in the near future, we hope—discussing biolinguistic matters. For example, we envision much of the backs and forths concerning material published in *Biolinguistics* (articles, reviews, forum pieces) to take place there. The URL for the *Biolinguistics* blog is http://biolingblog.blogspot.com.

We would also like to inform the larger biolinguistics community of the formation of a new Special Interest Group (SIG) on biolinguistics. SIGs are part of a new network mechanism recently initiated by the Linguistic Society of America (LSA) with the aim of sharing and exchanging information on a given topic, such as biolinguistics, among the SIG members (who also need to be members of the LSA). The LSA has not yet set up individual web pages for the SIGs, but this link has some information: http://www.lsadc.org/info/lsa-sig-in-formation.cfm.

The Biolinguistics SIG is coordinated by *Biolinguistics* co-editor Kleanthes Grohmann and Editorial Board member Bridget Samuels. The inaugural meeting took place at the last LSA Annual Meeting in Baltimore, and we're currently preparing a proposal to hold a Biolinguistics SIG Symposium at the next LSA Annual Meeting in Pittsburgh (6–9 January 2011), which non-members may also participate in. Additional information can be found at the new *Biolinguistics* blog: http://biolingblog.blogspot.com/2010/02/lsa-special-interest-group-on.html.

Lastly, some more exciting news: *Biolinguistics* just made another step towards true open access: After 3 years, we decided to get rid of the registration process for readers. We knew it was an unpopular feature of the journal (and subscribers encountered a number of technical difficulties as well), but this was the only way we had to track the interest in a new journal dedicated to an emerging field. With almost 2,500 subscribers (over 1,000 of whom have joined around the time of publication of the first *Biolinguistics* issue), the number of readers surpassed our expectations by far, and we now feel it is time to eliminate the registration step and make the journal fully Open-Access. From this issue on, there will be no more requirement for subscription (anyone stumbling across the journal webpage can read all materials collected here)—but registration is still required for authors; please read the Submissions section of the About menu item carefully where you also find a link to the current *Biolinguistics* stylesheet.

In this connection, we would like to point out that the new and revised stylesheet for *Biolinguistics* is accessible for everyone. It can be downloaded via http://www.biolinguistics.eu/index.php/biolinguistics/manager/files/BIOLIN GUISTICS_stylesheet.doc. Please bear in mind that all formatting and editing is done by ourselves (and since we don't have any professional training, it might appear somewhat idiosyncratic at times), so we are kindly asking everyone to please help us out by following the stylesheet to the best of your capabilities when submitting manuscripts; certainly manuscripts accepted for publication should follow the guidelines very closely.

Comments and feedback are, as always, appreciated. We hope that the new *Biolinguistics* blog will facilitate this kind of interaction more easily in the future, and we are looking forward to exchanges of ideas, suggestions, and more at this interface.

Kleanthes K. Grohmann　　　　　　*Cedric Boeckx*
University of Cyprus　　　　　　　*ICREA & Universitat Autònoma de Barcelona*
Department of English Studies　　　*Departament de Filologia Catalana*
P.O. Box 2053　　　　　　　　　　*7Edifici B, Campus UAB*
1678 Nicosia　　　　　　　　　　*08193 Bellaterra (Barcelona)*
Cyprus　　　　　　　　　　　　　*Spain*
kleanthi@ucy.ac.cy　　　　　　　*cedric.boeckx@uab.cat*

The Biological Nature of Human Language

Anna Maria Di Sciullo, Massimo Piattelli-Palmarini, Kenneth Wexler, Robert C. Berwick, Cedric Boeckx, Lyle Jenkins, Juan Uriagereka, Karin Stromswold, Lisa Lai-Shen Cheng, Heidi Harley, Andrew Wedel, James McGilvray, Elly van Gelderen & Thomas G. Bever

Biolinguistics aims to shed light on the specifically *biological* nature of human language, focusing on five foundational questions: (1) What are the properties of the language phenotype? (2) How does language ability grow and mature in individuals? (3) How is language put to use? (4) How is language implemented in the brain? (5) What evolutionary processes led to the emergence of language? These foundational questions are used here to frame a discussion of important issues in the study of language, exploring whether our linguistic capacity is the result of direct selective pressure or due to developmental or biophysical constraints, and assessing whether the neural/computational components entering into language are unique to human language or shared with other cognitive systems, leading to a discussion of advances in theoretical linguistics, psycholinguistics, comparative animal behavior and psychology, genetics/genomics, disciplines that can now place these longstanding questions in a new light, while raising challenges for future research.

Keywords: language development and genetics; linguistic competence in a comparative ethological context; linguistic theory; neurology of language and the genome

1.   Introduction: Background and Overview

This short article brings together for biologists and linguists recent hypotheses, studies, and results on the human faculty of language stemming from the Biolinguistic Program. It is not intended to cover in full all current research relating biology and language; nor could it hope to do so. Rather, the specific topics discussed here may be taken as illustrating just some of the current lively

research in this field.

The modern Biolinguistic Program initially grew out of collaboration between biologists and linguists in the late 1950s and early 1960s, initially Eric Lenneberg and Noam Chomsky, and later Salvador Luria among others, epitomized in Lenneberg's *Biological Foundations of Language* (Lenneberg 1967). A 1974 international meeting at MIT called *Biolinguistics* by its organizer, Massimo Piattelli-Palmarini, propelled it further. More recently, books and papers surveying the field and proposing new directions have been published (e.g., Jenkins 2000, 2004; but for a very different approach, see Givón 2002); and conferences have been organized worldwide, coupled with the 2006 launch of this journal, *Biolinguistics*, and the foundation of an international network on biolinguistics in 2007 (www.biolinguistics.uqam.ca; for an overview, see the forum contribution to this issue, Di Sciullo 2010).

A variety of contemporary theories and research programs provide a way to relate biology and language. For example, the 'Minimalist Program' (Chomsky 1995 and related works[1]) provides a methodology to explain linguistic phenomena with minimal theoretical apparatus, attempting to make contact with biology, physics, psychology, and computational neuroscience.[2] The Minimalist Program in the broad sense, including the insights discovered in earlier work, such as the Principles–and–Parameters model (Chomsky 1981) and the formal approach to syntactic structure (Chomsky 1956), has been very successful in shedding light on a number of the questions raised by Biolinguistics. This is not to say, for example, that Minimalism predicts the *FOXP2* gene in language or the organization of the language areas in the brain.[3] Given our current understanding, no research program or theory can do this. For the moment, it suffices that a linguistic research program should suggest productive avenues to explore in order to illuminate the biological nature of language.

The Biolinguistic Program proceeds by trying to answer (classical) questions about the mechanism, the development, and the evolution of language. *What is knowledge of language? How does language develop in the child? How does language evolve in the species?* Each can be studied at different levels. Knowledge of language can be studied at an abstract, top level, by positing a 'faculty of language' that includes a generative grammar with various properties, including recursion, structure-dependence, symmetrical and asymmetrical properties, and the like. Knowledge of language can also be studied at the neural level, using tools of brain area mapping (e.g., Broca/Wernicke areas, noun/verb regions,

---

[1]   See, e.g. (in alphabetical order), Boeckx (2006, 2008), Chomsky (2000, 2001, 2005, 2009a, 2009b), Di Sciullo (2003a, 2003b, 2005, in press), Hinzen (in press, to appear), Lasnik (1999, 2002, 2003, in press a, b), Martin & Uriagereka (2000), McGilvray (2005, 2006), Uriagereka (1998, 1999), and Uriagereka & Boeckx (2007).

[2]   See, among other works (in alphabetical order again), Belletti & Rizzi (2002), Berwick (in press), Berwick & Chomsky (in press), Berwick & Weinberg (1986), Boeckx & Piattelli-Palmarini (2005), and Di Sciullo & Boeckx (in press).

[3]   We will use the following nomenclature for FOXP2 (see also http://biology.pomona. edu/fox): "Briefly, nucleotide sequences are italicized whereas proteins are not. Human forms are capitalized (e.g. FOXP2 protein), murine forms are in lowercase (e.g. Foxp2), and those of other species, such as the zebra finch, are in uppercase and lowercase (e.g. FoxP2)" (Teramitsu *et al.* 2004: 3152).

etc.), imaging, probes, and so forth. Ultimately, we want to link grammars to the brain, but at this stage we do not expect to be able to predict properties of neural organization from properties of grammar. But if we find that, for example, certain areas are active for verbs and others for nouns, then this is consistent with a grammatical theory with a verb/noun distinction, but not with one that has no such distinction. Other topics such as brain lateralization and handedness can also be studied at multiple levels (Bever *et al.* 1989, Geschwind & Galaburda 1985). Similar remarks apply to sign languages.

Similarly, the Biolinguistic Program can study the development of language at an abstract level, positing a Universal Grammar to account for both universal properties and language variation (say, by adopting parameters along the lines of Chomsky's Principles–and–Parameters program, together with a probabilistic model as proposed by Yang (2002), or Kayne's (2005) micro-parameters in comparative linguistics. Alternatively, one can study this area more concretely by looking at the developmental trajectories of actual children (Wexler 2003). In just the same way, critical periods can be studied abstractly (Stromswold 2005) or concretely by looking at genetic programs in other critical period systems. In some cases one can even investigate at the level of single genes, as with *FOXP2*, where one can link the abstract patient behavior (verbal dyspraxia) with genetic mutations (Gopnik 1990, Vargha-Khadem *et al.* 1995). We can do the same with other genetic disorders—DNA duplication, chromosome disorders, language in Williams Syndrome, dyslexia, etc. Finally, the Biolinguistics Program can address the question of the evolution of language (as well as language change) at the abstract level by computer simulation (Niyogi & Berwick 1997, 2009, Nowak *et al.* 2002, Niyogi 2006,) or via cross-species behavioral studies (starlings, cotton-top tamarins, etc.). Again at a more concrete level, one can carry out cross-species comparisons of *FOXP2* and of other genes affecting language. Finally, questions as to why the language faculty has certain properties and not others can be linked to all three levels above, and might include memory constraints on parsing, economy conditions, (a)symmetry, or—more concretely—wiring minimization.

Crucially, the Biolinguistic Program explicitly factors the 'faculty of language' into three components consonant with biology generally—language's genetic endowment, 'Universal Grammar', environmental experience, and biophysical principles that are language-independent (Chomsky 2005, and related works).[4] It connects historically with a line of inquiry attentive to the 'laws of form', that is, to the topological, computational and self-organizational invariants of life forms, going back to D'Arcy Thompson and Turing.

A major aim of the Biolinguistic Program has been to explain why Universal Grammar, extracted from commonalities across languages, is what it is and not something else. This basic question leads to an investigation of what the genome specifies that is particular to language, and raises the possibility that this genetic endowment specifies only a few, basic computational features. This approach has already led to two empirical benefits for research. First, by specifying

---

[4]   See, again in alphabetical order, Berwick (in press), Chomsky (2009a, 2009b), Di Sciullo (2007, in press), Jenkins (in press), Lasnik (in press), Medeiros (2008), Piattelli-Palmarini & Uriagereka (2004, 2005, 2008, in press), Piattelli-Palmarini *et al.* (2009), Uriagereka (in press).

a relatively compact set of computational principles, it enables a more direct link to animal behavior and evolutionary analysis than previously considered. This broader comparative approach enables a more systematic exploration of the components of the language faculty that are shared with other animals and other domains of human knowledge (the 'broad' faculty of language) and those that are unique to humans and language (the 'narrow' faculty of language) (Hauser *et al.* 2002). Second, it is now possible to integrate our understanding of how cognitive systems of learning and use constrain language universals with a general theory that spans biology and formal linguistics (e.g., Miller & Chomsky 1963, Bever 1970, Wexler & Culicover 1980).

We begin by providing a brief synopsis of minimalist proposals about the faculty of language and their import for Biolinguistics. We then turn to four sections that showcase discoveries that are consonant with minimalist proposals and that contribute to our understanding of the biological basis of language. First, we discuss recent findings on the acquisition of formal grammars by monkeys and other non-human species. We note that while there is also extensive work in aphasia and brain imaging, we have chosen to single out these cross-species analyses because they aim to provide experimental results on the human specific properties of the faculty of language. Second, in the area of language acquisition and child language disorders, we focus on tense and finiteness. Here too, whereas much current work is devoted to other linguistic aspects, such as difficulties in sentence structure and movement, the results of the study on tense and finiteness relate to the hypothesized properties of the faculty of language, including the asymmetry of (morpho-syntactic) features/projections. These studies also lead to promising genetic correlates. Third, we review recent advances in the genetics of language, with special emphasis on *FOXP2*. Finally, we turn to results on naming and dextrality/sinistrality (handedness) as they relate to the minimal assumptions about the faculty of language and lead to further inquiry in the relation between language and biology/genetics. Once again, there has been extensive work on anomia and its lexical basis, as well as syntactic impairments in aphasia, that are relevant to the discussion on whether syntax and the lexicon have different neurological foundations and which brain areas might be correlated with certain linguistic functions, but we have chosen to high-light the particular case of handedness as a 'case study'. We conclude with a discussion of the promise of the Biolinguistic Program, as well as challenges for the future.

## 2.   Theoretical Linguistics and Biolinguistics

The configuration in (1) represents a minimalist proposal for the skeletal architecture of the human faculty of language, with a schema familiar to biologists as a framework for modeling other complex processes. It is the bare bones of our core linguistic apparatus.

(1)   NS
        CI   SM
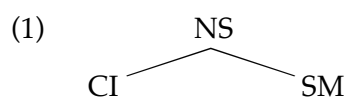
This simple machinery produces a potential infinity of corresponding pairs of sounds and meanings. These paired instruction/information sets are passed on to and then interpreted by separate systems, external to the core computational system. The first is the Sensory-Motor system (SM): It satisfies the biological requirements imposed on the production and the perception of language (auditory in spoken languages, visual in sign languages). The biological requirements correspond to the structure and function of the auditory system in perception and to those of the vocal tract and articulatory apparatus in production. The second is the Conceptual-Intentional system (CI): It satisfies the biological requirements involved in interpretation, reasoning and inference, internal to the brain's other cognitive functions (Chomsky 1995, 2005). Adopting this division represented in (1) reduces to a minimum the parts of this architecture that are particular to human language, as opposed to more general kinds of biological computations. Crucially, like all biological machinery, (1) has only finite computational power, but must still produce an infinity of possible sound-meaning pairs, in Humboldt's famous phrase making "infinite use of finite means". This is the basic function of the narrow faculty of language, solving its core biological 'design problem', producing what is called 'narrow syntax' (NS). Conjecturing that (1) may be reduced to such a minimum is a strong claim about what the human genome specifies as the unique core of universal grammar. If this hypothesis is correct, then this language-specific genomic component is optimal in the sense that it is sufficient to solve the core system 'design' problem and no more. To draw an analogy to an example familiar to biologists, just as the human genome (with only about 24,000 genes) cannot and need not encode the precise neuron-to-neuron wiring of the brain (Cherniak *et al.* 2004), NS need not specify that spoken language must be produced according to the word-order rules of each particular language. There are solid grounds to suppose that word-order constraints are implicitly imposed by a language community from a biologically limited set of possibilities and by physical constraints on the articulatory system—we cannot say two things at once. In this way, what must be specified by the genome for NS is held to a bare minimum.

What remains? Sentences are obviously quite different from mere lists of words. While linguistic researchers often disagree on the details, they largely concur that the constituents of sentences are hierarchically structured phrases, themselves consisting of smaller units, commonly referred to as words, with their own internal morphological structure. In current computational theory, there is but one way to construct such an infinite variety of hierarchical structures using a finite set of elements, and that is via a recursive operator that combines parts together into new, larger wholes: Larger components are built from smaller parts. The elements of a sentence are combined by a local operation that connects components immediately adjacent in the hierarchy. The structures in (2) outline the simplest form of this operation. Characteristically, each sub-hierarchy is composed of a tree with a category node that dominates two categorized branches. A sub-hierarchy can be merged with another when the category at the top of one sub-tree matches one of the bottom branches of another, as shown. This operation can then apply again to its own output, building up more complex hierarchies, yielding an infinity of outcomes—sentences may contain other sentences, and

these in turn other sentences. In the Minimalist Program, this constructive operation is called 'Merge': This term is intuitively based on the fact that it connects two 'sisters' at the same hierarchical level, and then these together, as a unit, to the immediate higher node in the hierarchy, and then again and again to produce a more complex hierarchy. Many other linguistic traditions adopt principles similar in spirit, since there must be some means of solving the combinatory problem such that a finite system yields a potentially infinite output.[5]

(2)   a.   X   b.   Z   c.   X

              Y   Z   P   Q   Y   Z

                                                                                        P   Q
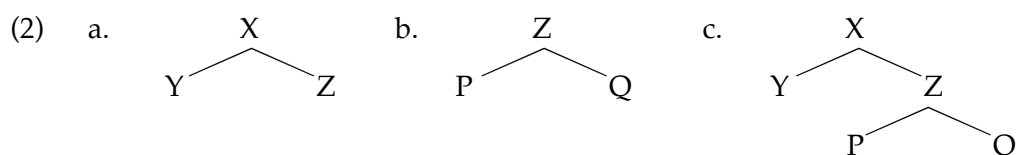
Thus a central goal of the Biolinguistic Program is to characterize the biological properties of this recursive operator, beginning with its abstract properties, ultimately arriving at its concrete biological instantiation.[6] To be sure, at this stage we know very little about how recursive computation is actually implemented in the brain. But at an abstract level, we know that Merge must have properties beyond recursion, since recursion alone is a generic property of any infinite output system, including mathematics and music. In particular, Merge must also be asymmetric in singling out one of the two initial elements (sisters) it operates on instead of another, for example, either always the left member of a tree like that in (2a), Y, or the right member, Z. Otherwise, all linguistic elements would be treated identically and there would be no differentiated structure in words, phrases, sentences, or anywhere else. In addition, Merge must be asymmetric in selecting the elements to which it applies. The sets of features of these elements must be in a superset/subset relation otherwise it would not be possible to satisfy their morpho-syntactic features. Thus, verbal inflection features, such as Tense, combine with verbal elements/projections; they do not combine with nominal elements/projections. The feature asymmetry stemming from the dynamics of Merge could be specific to the human language faculty. If so we would not expect to find manifestations of feature asymmetry in non-human primates and in other animals (see section 3). We would however expect the ability to compute feature asymmetry to be affected in specific language impairment (see section 5). It might be the case that the asymmetry of Merge and the hierarchical structures it derives is rooted in biology.

This is reminiscent of the way that asymmetry in developmental gradients derives structure in embryogenesis, or asymmetry in cell–cell adhesion or contractility generates mechanical forces that can be deployed combinatorially, developing a broad range of epithelial tissue types in morphogenesis (Montell 2008).[7] This is also conceptually and computationally similar to the current

---

[5]   This problem is also addressed in frameworks other than Minimalism, such as Categorial Grammar, going back to Ajdukiewicz (1935), Bar-Hillel (1953), and Lambek's (1958) seminal work, modeling the combinatory possibilities of the syntax of human languages.

[6]   See Hauser *et al.* (2002) on recursion, as a property of the language faculty, and Fitch (2010) on the importance of this notion for biolinguistics.

[7]   See also Boeckx (2006), Chomsky (1995, 2005), Di Sciullo (2003a, 2003b, 2005, in press), Di Sciullo & Isac (2008a, 2008b), Gelderen (2004, 2008), Hinzen (in press), Kayne (1994, 2002, in

inquiry into evolution and development (the 'evo–devo' theory of evolution) which underlies recent biological explanations of how the same battery of genes engenders a variety of life forms by small changes in the regulation of their timing and rate of expression. As with the morphological segmentation of body plans, the minute factors that produce repeated breaks in symmetry mold language's shape in sometimes surprising ways. This is linguistic science's own twist on the combinatorial power of simple units, explaining the diversity of languages, analogous to the biological processes that produce, in Darwin's famous expression, "endless forms most beautiful" (Raff 1996, Jenkins 2000, Carroll *et al*. 2001, Carroll 2005).

3.   Linguistic Competence in a Comparative, Ethological Context

Recent work on the evolution of language has turned from looking strictly at communication, to exploring the similarities and differences between humans and animals with respect to computational competence. This new angle has opened the door to exploring the capacity of animals to extract artificial grammars that represent the building blocks of linguistic syntax. This work suggests the novel possibility that humans share with animals some of this core foundation, but uniquely evolved the capacity to interface syntactic structures with semantic and phonological representations. In this way, experimental work with non-human primates provides data that can be used to identify the commonalities among human and animal languages.

Several studies build on this idea, using the formal approach to syntactic structures laid out by Chomsky (1956)—the formal hierarchy of rules and regularities—together with experimental work in artificial language learning. This research isolates specific kinds of linguistic computation that are most relevant to acquisition, including problems of segmentation, the extraction of algebraic rules, the relationship between types and tokens, and the relationships between abstract variables. For example, infants can use transitional probabilities to segment a continuous stream of speech; corresponding studies of rats and tamarin monkeys have provided parallel evidence. Adult humans and infants can extract abstract rules of the form AAB, ABA, and ABB; corresponding studies with rats and monkeys show similar abilities. Tamarin monkeys and starlings can recognize the strings generated by a grammar that places symbols in a simple line like beads on a string (patterns in the form $(AB)^n$). It has been argued that starlings reveal suggestive abilities—after a training that consists of several thousands of repetitions—at recognizing the strings of the next order of grammatical complexity generated by a grammar that 'nests' or embeds matching words or symbols hierarchically (patterns in the form (AB), (A(AB)B), (A(A(AB)B)B),…, A^nB^n) (Fitch & Hauser 2004, Gentner *et al*. 2006). More recently, such results have been called into question by more careful experimental tests with zebra finches suggesting that non-nesting rules alone suffice to account for

_____

press), Langendoen (2003), Moro (2000, 2008), and Zwart (2006), for example, on the role of asymmetry and symmetry-breaking in the derivation of linguistic expressions.

the results (van Heijningen *et al.* 2009). Additional evidence of successful discrimination among tokens based on a particular rule leaves open alternative solution strategies. For example, strings of the form A^nB^n could be recognized by computing the hierarchy and embeddedness of AB pairs or simply by counting and checking for an equal number of A's and B's. While preliminary results are already available (see Figure 1, Friederici 2009), further work is necessary to understand the formal properties of animal computations and how they differ from the ones derived by human computation (Merge).[8] More interesting artificial language paradigms could be tested to see whether tamarin monkeys and other animals can generate/recognize asymmetric relations, be they featural or hierarchical. For example, while humans compute indirect recursion, for example, *the queen's garden's roses*, *the cat on the tree in the garden*, experimental tests with animals using artificial language paradigms expressing left and right indirect recursion asymmetries, for example, AC BC D, where there is a different number of A's and B's and where C would stand for a functional category/ projection such as the possessive '*s* or a preposition in the previous examples from English, could be telling about the kind of complexity and asymmetric relations that can be computed by animals. There is much to be done in this area, as, to start with, it is unclear whether, given a set of elements {A, B, C}, animals may generate/ recognize asymmetric relations, for example, AB BC AC or CB, CA BA, and so on, as opposed to symmetric ones, for example, AB BA AC CA BC CB.[9] In any case, and notwithstanding the difficulties encountered to test the underlying computations in animals as well as in humans, this kind of work is likely to shed light on the species-specific computational machinery behind human language, by exploring the similarities and the differences with animal computation.

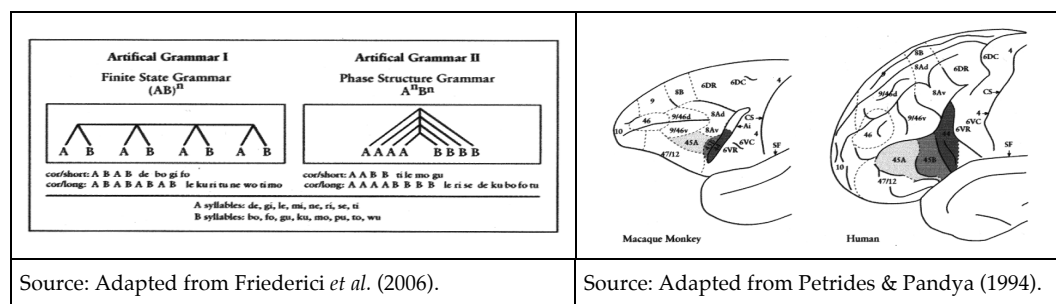| Source: Adapted from Friederici *et al.* (2006). | Source: Adapted from Petrides & Pandya (1994). |

Figure 1: A difference in the kind of structures generated by Finite State Grammars vs. Phrase Structure Grammars [left]; brain imaging of the human brain and the macaque brain show differences in the size of Broca's areas (BA44, BA45) [right]

---

8   The role of Broca's areas in processing syntactic structure has been the topic of extensive research in aphasia and brain imaging studies for more than three decades (Grodzinsky & Santi 2008). Recent results from brain imaging (Friederici 2009) provide further evidence of the role of these areas (BA44, BA45) in syntactic (hierarchical) processing. See also Endress, Cahill *et al.* (2009) and Endress, Carden *et al.* (2009) on the processing of sequence peripheral positions in apes, and Mody & Fitch (submitted) on artificial grammar learning by humans, extending the standard paradigm to mildly context sensitive grammars.

9   See Huber *et al.* (1999) on the limits of symmetry/asymmetry visual pattern detection in pigeons, and Swaddle *et al.* (2008) on the limits of visual detection of fluctuating asymmetry in starlings.

Furthermore, work on the biological basis of vocal learning in animals, such as passerine songbirds (e.g., Jarvis 2004, 2006), has contributed to our understanding of the computational machinery behind human language acquisition. This research brings an important new perspective to our understanding of the origins and the evolution of language and its relationship to animal communication: a key shift from looking for communicative similarities or overlap to looking for common or shared competence in certain kinds of computation.

There are, however, differences between human and animal communication. The following illustrate some of these differences stemming from field studies in behavioral psychology. Studies of natural behavior have largely focused on two core properties of language: referentiality (i.e. symbol–to–world relations) and combinatorics (i.e. syntax). Field biologists have documented in vervet monkeys different specific calls in different specific contexts, while playback of those calls elicits corresponding behaviorally appropriate responses: In response to hearing the playback of an alarm call, vervet monkeys respond as if a predator was nearby. Such examples have been used to argue that animals produce vocalizations that are like our words. That is, they are formed on the basis of an arbitrary association between sound and meaning, such that each sound can trig-ger a representation of the target object or action—they are *functionally referential* (Seyfarth *et al.* 2005, Zuberbühler 2003). Upon closer inspection, however, the parallels between these sounds and our words are un-impressive. For example, if these animals had really evolved the capacity for referential expression of all concepts available to them, then it is unlikely that their lexicon would be limited to 10-20 'words'. Similarly, if their referential capacity was like ours, then animals should be able to refer to a wide range of objects and events, real or remote, past, future and imagined. But they cannot: All of the putatively referential calls they produce map onto a narrow range of actually present objects or events, even though these animals confront a wide range of social and ecological situations that are, from a functional perspective, *worthy* of comment.

With respect to combinatorial syntax, two sets of findings have entered the discussion. On the one hand, songbirds (e.g., Hailman & Ficken 1987) and whales (e.g., Suzuki *et al.* 2006) have spectacular abilities to combine their vocalization elements iteratively to create new forms in ways that could be thought of as derived by Merge. However, this combinatorial facility is independent of their conceptual system. When a songbird or whale recombines notes to produce new songs, the meaning of the signal remains the same: a sound that identifies the individual, its population or group, often to attract a mate and fend off competition. In contrast, recent studies of non-human primates suggest that when individuals combine vocalizations, these new strings are different from the meaning of each element in the string (e.g., Arnold & Zuberbühler 2006). Here too the syntactic facility is strikingly different from human language. Specifically, it operates over only two sound types and two at a time only; these types are not abstract categories such as noun and verb, or morpho-syntactic features. Thus, this primate capacity is limited to a single combination, far from the 'infinite use of finite means' that is the hallmark of human language.

Studies of animals trained to use human-created symbols support this

account. Animals clearly can be taught to use many symbols in a functional context, and may be able to combine these symbols. Apes can acquire several hundred symbols, sometimes apparently combined into new, meaningful strings. Yet, the rich induction that every two-year-old child makes during word learning never arises, leaving the animals with a highly fixed lexicon and children with an open-ended and explosive one. Furthermore, only human children combine words together, combining affixes and roots to form new words. Animals in these training situations are frozen at what language acquisition researchers dub the one-word stage (Terrace *et al.* 1979).

Yet, animals sometimes acquire some understanding of spoken language. Thus, one dog showed evidence of 'fast mapping', that is, novel word learning from minimal data (Kaminski *et al.* 2004). Studies of bonobos exposed to language suggest that individuals understand some aspects of word order and that words are classified into categories such as object, action and location (Engor *et al.* 2004). Experimental studies of language comprehension in animals are more promising than those of production, suggesting that a fundamental bottleneck in the evolution of language was the connection between linguistic computation and its externalization in linguistically meaningful structures.

4.   The Genetics and Evolution of Language: Promises and Pitfalls of Modern Genomics

Until recently, very little has been learned about language from animal studies or from comparative psychology or biology. For many years, a single *idée fixe* held sway, namely, the species specificity of language—its uniqueness to humans—as well as its phenotypic uniformity within individuals. This was entirely unsatisfactory from an evolutionary standpoint, because ever since Darwin the *sine qua non* of evolutionary analysis has been both the comparative method and explicit acknowledgement of individual variation. Within the past five years, however, this picture has completely changed: Variation and concomitant evolutionary analysis is taken seriously as a biological aspect of language. It is now wellestablished that genes affect speech and language in individuals and there are now many demonstrable associations between inter-individual differences in genetic makeup and inter-individual differences in speech and language abilities. Perhaps the best-known is the recent *FOXP2* genomic analysis, which can be carried all the way from nucleotide variation to protein variation to embryonic development to brain function, as well as deployed for comparisons against other species including extinct hominid species like Neanderthals. But this is not an isolated example. There now exist behavioural genetic studies dissecting heritability, detailed language disorders, and language variation, all fitting into the familiar biological analysis of a complex, polygenic trait. In contrast to the classical results focusing on abnormal language and aphasias, this recent research has found that even in typical language development there appears to be genetically linked variation, some of it highly specific. Extremely detailed patterns of syntactic development, once thought to be the sole province of academic linguists, such as finiteness (tense/lack of tense) have been thrust into the biological

arena and are now known to be mostly inherited and their genetic source distinct from that of another inherited trait, phonological working memory, with established links to autism and Williams syndrome. Rounding out the connections between genetics and biology, in the case of language variation and its geographic distribution over time and space, we can now employ dynamical system analysis and computer simulations just as in population biology, accounting for the known distribution and trajectories of language variation and change, parallel to biological evolutionary ecological analysis. Taken together, such work has opened the door to how genetics might bear on the development and inheritance of particular properties of language, not just language as a whole. The following paragraphs outline some advances in our knowledge of the connections between language and genetics, as well as some of the pitfalls.

Genetics and evolutionary analysis have worked productively using the comparative approach: Looking to other species to establish parallel similarities and differences. This poses special challenges in the case of language, since no other animal species seems to fully possess the human language phenotype. What then can classical and modern genomics tell us about language's genetic and evolutionary properties? Some simple questions can be answered immediately. Classical heritability studies establish beyond a doubt that there are language components that are 'innate', that is, that have a significant genetic component, since estimates for the additive variance attributable to genes ranges between 0.4–0.6 (Stromswold 2005), roughly comparable to hair color. Likewise, as a complex behavioral phenotype, language is clearly polygenic. However, language is also labile in a special way that hair color is not, since the particular language a child speaks depends critically on the language of their caretakers, who need not be his/her genetic parents. To go beyond this broad characterization, recent research has embarked on a more precise genetic and functional dissection of language, including for the first time the isolation of at least some genomic elements that appear to have analogs in other species such as songbirds and mice. We outline just two new directions here.

One approach has opted for a more careful and systematic heritability analysis of the subcomponents of language via twin studies, revealing a closer genetic overlap between syntax and sound-structure (phonological) abilities and fine motor control than syntax and the lexicon (Stromswold, in press). Going forward, by pushing conventional quantitative trait loci (QTL) methods to their limits, we may be able to characterize even more finely the differences between conventional linguistic categories such as syntax, phonology, the lexicon, and semantics (Stromswold, in press; see also Stromswold 2001, 2006, 2008). More speculatively, given the availability of high-throughput whole-organism genome scans, we may be able to supersede conventional QTL, as has been done in other polygenic cases like Alzheimer's syndrome, since it has now become feasible to test all of an organism's genes directly for association with some trait of interest. In the present case, what this means is that for some very particular syntactic ability, we may be able to identify a candidate gene set correlated to this trait.

Current research has embarked on a more precise genetic and functional dissection of language. This was recently sparked by the discovery of a Mendelian point mutation linked to a rare language disorder across several gen-

erations in a single British family. While the exact phenotypic manifestation of this disorder remains open to debate, many agree that the disruption results in so-called 'verbal dyspraxia', including a general inability to orchestrate coordinated mouth movements. However, a range of language and non language abilities are apparently impacted, including written language as well as phonological working memory (Fisher *et al.* 2003, Watkins *et al.* 2002). Researchers eventually identified, cloned and sequenced the gene in question, *FOXP2*, making it the first discovered gene causally linked to language (Hurst *et al.* 1990, Vargha-Khadem *et al.* 1995, Fisher *et al.* 1998, Lai *et al.* 2000, Lai *et al.* 2001, Vargha-Khadem *et al.* 2005). It is one of a subclass of an evolutionarily well-known family of 'forkhead box' transcription factor genes whose protein product interacts with DNA, thus regulating other genes.

The discovery of the *FOXP2* gene was initially greeted with great enthusiasm. It was the first specific genetic window into human language development and evolution, and offered hope of linking human and animal studies into one, with the possibility of testing an animal model. However, subsequent analysis strongly suggested that the protein transcripts of this gene, FOXP2, might not be implicated directly in the central computational aspects of language. Individuals with the language deficit have sequential control motor deficits not limited to language syntax (Haesler *et al.* 2004, Hauser & Bever 2008). Its role in other species suggests that it underlies a general vertebrate sensorimotor system for fine motor control and 'higher level' sequential movement planning. For example, FOXP2's homologues have been implicated in the production, learning and perception of songs in oscine songbirds (Webb & Zhang 2005, White *et al.* 2006, Teramitsu & White 2006, Haesler *et al.* 2004) with FoxP2 expression boosted during the seasonally plastic song-learning periods. Further, FoxP2 suppression in the zebra finch song-learning brain area leads to inaccurate song imitation and acquisition by juvenile learners (Haesler *et al.* 2007). FoxP2 has also evidently been undergoing recent rapid evolution perhaps linked to the vocal-learning of certain bats (Li *et al.* 2007). It may even be related to the proper development of ultrasonic calls in mouse neonates (Shu *et al.* 2005, Fujita *et al.* 2008, Groszer *et al.* 2008,). Thus, *FOXP2*'s role in human language may underlie part of the machinery that builds the sensorimotor system for fluent speech. Neural activity studies in the affected family (Vernes *et al.* 2006) had promoted the view that FOXP2 affects the regions involved in general planning of fine motor output, sensorimotor integration, and multimodal sensory processing, as opposed to circuitry controlling mouth and lower face movements. Genetically engineered mice exhibit altered motor control learning (French *et al.* 2007, Groszer *et al.* 2008), which seems to support the initial consensus that FOXP2 is critical for learned motor skills rather than language *per se*. However, Teramitsu & White (2006) showed that the expression of the gene is different depending on whether the bird's song is directed or undirected. The motor control is the same, but the expression of the FoxP2 gene is different (there is down-regulation in the undirected singing vs. slight up-regulation in the directed singing).

The *FOXP2* discovery also jump-started modern evolutionary Biolinguistics. There are just two functional amino acid differences between *FOXP2* and its variant in chimpanzees, plus one additional difference with the variant in

mice. The two human–nonhuman primate differences have been interpreted as the effect of accelerated evolutionary change in the 4.6–6.2 million years that separate us from the chimpanzee, possibly with a faster rate within the last 50–100,000 years, and have been posited as the target of positive natural selection, perhaps concomitant with language emergence (Enard *et al.* 2002). Not surprisingly, the causal relationships between these two changes and language remain unclear. Further, cautionary notes must be sounded regarding the contributions of selective processes as opposed to non-selective processes such as biased gene conversion (Berglund *et al.* 2009, Duret 2009, Hodgkinson *et al.* 2009).

Our understanding of *FOXP2* has itself evolved, unsurprisingly since *FOXP2* is one of the largest and most complex regulatory genes known. The initially clear-cut, one-gene–one-behavioral phenotype FoxP2 picture has been replaced with a much more nuanced 'molecular network' systems view (Fisher & Scharff 2009) in which many 'downstream' cognitive systems might be affected by the *FOXP2* gene. Taking into consideration the words/non-word repetition task deficits that the KE family members exhibit (Watkins *et al.* 2002), some have suggested a relation between the regulation of the expression of *FOXP2* and procedural or working memory (Bosman *et al.* 2004, Ullman & Pierpont 2005). Building on this general approach, although focusing on the specific memory specifications that phrasal manipulation requires (basically, a push-down automaton), Piattelli-Palmarini & Uriagereka (2005) propose that such matters are relevant to architectural computational concerns in competence systems. On the basis of reduced activity in Broca's area, still others suggest a possible link with mirror neurons (Corballis 2004). Furthermore, the interesting parallel with the expression of the gene in songbirds has suggested a possible role for FOXP2 in the linearization of complex hierarchical structures into a linear sequence, as well as its reconfiguration, upon successful processing into the original internal structure (Piattelli-Palmarini & Uriagereka, in press).

Moreover, Vernes *et al.* (2008) found that FOXP2 binds to and dramatically down-regulates CNTNAP2, a gene that encodes a neurexin and is expressed in the developing human cortex. On analyzing CNTNAP2 polymorphisms in children with typical specific language impairment, they detected significant quantitative associations with nonsense-word repetition, a heritable behavioral marker of this disorder. Intriguingly, this region coincides with one associated with language delays in children with autism (Piattelli-Palmarini & Uriagereka 2005, in press). Currently there is no way to decide among these various possibilities, since we are still uncovering important new details about the basic genetics of this complicated system. (For example, recently it has been found that *FOXP2* is itself regulated by a host of so-called 'small RNAi' or sRNAi molecules (Friedman *et al.* 2009) and a new *FOXP2* transcription 'start site' was also discovered (Schroeder & Myers 2008).) So, we have some distance to go in understanding the complete FoxP2 picture. Nevertheless, all this is an advance, not a retreat, since *FOXP2* will undoubtedly serve as a role model for future genomics research about language. While it remains an unfinished task to identify the final causal links to language impairments, the discovery and analysis of this gene and the genes that it regulates have served as an extremely useful example of how to unravel via genomics the complex phenotype that is human language.

5. Language Development and Genetics: The Case of Tense

Within the last decade, there have also been major advances in our under-
standing of language development, especially the precision, replicability, and
meaningfulness of particular results. For the first time we are in a position to
state surprising and non-trivial results about the nature of language develop-
ment. The traditional idea was a false-to-fact idealization about instantaneous
development, as if information was available to the child at a single point in time.
Current practice and results turn this original notion on its head, rendering the
concept of child's biological development central rather than peripheral. In this
sense, the classic work on language and biology—Lenneberg's (1967) *Biological
Foundations of Language*—was right, but appeared still-born. We wish to point out
some new results and sketch how they fit into a broader biology. New methods
make this possible: The level of quantitative precision about computationally
precise developmental linguistic behaviors has raised the descriptive bar by an
order of magnitude compared to a decade ago. In some cases, the results bear a
striking resemblance to scientific laws. Further, it is a truism that human
language capacity combines learned and genetically-transmitted abilities, and we
must therefore take seriously the interaction of genetics and language. For the
most part, this is currently accomplished via studies of deviant populations. In
the past this had proved difficult because we lacked the requisite information
about typical development. Given major advances in the field, we can now study
impaired populations, and a surprising number of regular results are beginning
to be uncovered across a range of linguistic and cognitive deficiencies. These
findings seem to hold great promise in the quest for genetic understanding; for
instance, for the first time, both behavioral genetic and linkage studies are being
carried out in the normally developing child as well as in children with selective
impairments. We will review these—what is already known and what can be
accomplished in the (near?) future. These new methods involve an exquisitely
precise understanding as to how linguistic representations develop, and how to
relate these to concrete deficiencies.

Language development may be influenced by general learning procedures,
Bayesian inference serving as a prime instance currently under investigation in
many research programs. In that light, a central question for Biolinguistics is
which, if any, specialized structures have evolved in the service of such processes
for language in particular, which have developed for cognition in general, and
which are only accidentally affected by such inferential mechanisms in indivi-
duals. In the last 20 years, there has been intense study of the early computational
system of language in children. The results show that very young children 'learn'
basic properties unique to their language much faster than unguided learning
models could predict. Such phenomena have motivated linguists to postulate a
strong genetic component that pre-figures possible languages in the child's mind,
as is typical of species-specific 'learning' in general. Furthermore, the ability to
acquire language as a first language decays rapidly after puberty (Lenneberg
1967). Research investigating the cognitive and neural bases of language acqui-
sition addresses the following questions about this phenomenon: How abrupt is
this diminution in language learning ability? Is it an averaging effect? Are all

aspects of language affected equally and at the same time?

In the following sections, we discuss molecular genetic studies of second language learning, as well as behavioral genetic studies (twin studies, family aggregation) of second language learning, shedding light on a deeper understanding of why there is a critical period for language acquisition. Current research explores the impact of sex hormone levels on language learning using the variability in onset and duration of puberty in a variety of clinical and normal populations. The investigation of the relationship between genetics and the so-called critical period, for example, molecular genetic studies of second language learning, as well as genetic studies of second language learning, permits a far deeper understanding of why there is a critical period, how it is related to learnability theory, language processing, and language evolution, as well as unraveling the relationship between language learning and neural activity and its structure.

5.1.   *Very Early Clause Structure Variation*

Although basic properties of language may be genetically determined, languages differ in some experience-determined ways. For example, unlike English, Dutch and German are examples of languages that are called 'verb-second' (V2): The verb in the main clause always comes second in the sentence (otherwise it is always at the end of the clause) and is tensed (e.g., present or past), regardless of what comes first. Any other part of the clause may be in the 1st position. The German sentence *Das Buch hat Johann* 'the book has John' (meaning 'John has the book') shows the verb *hat* 'has' in 2nd position, though *das Buch* is not the subject. Only verbs that show 'tense' can enter the V2 (Tense) position. In adult language, all main clauses have a tensed verb. Yet, young children very often omit tense, producing an 'infinitival' verb (an 'Optional Infinitive') like *Johann das Buch haben* 'John the book (to) have' (Wexler 1993, Poeppel & Wexler 1993, Haegeman 1994, Guasti 2002). Why is this? If children know that their language is V2, they should place tensed verbs in 2nd position and only verbs without tense in final position. Many studies of the sentences spontaneously pronounced by young children speaking V2 languages show that this prediction is borne out almost perfectly (Poeppel & Wexler 1993, Haegeman 1994, Wexler *et al.* 2004). For example, Wexler *et al.* (2004), studying 2,590 utterances of Dutch children between 1;7 and 3;6 years of age, found that about 1% of the utterances violated the prediction. This developmental pattern holds in all V2 languages studied so far. Children learn the V2 pattern very early and well, suggesting that children are brilliantly plastic with respect to certain variable features that differentiate languages. They do not, however, retain this plasticity.

5.2.   *Why Untensed Verbs?*

Why do young children use so many verbs without tense in simple clauses, which are violations in the adult language? The systematic nature of their patterns shows that this is not a trial–and–error phenomenon. One hypothesis (Wexler 1998, 2003) is that children are subject to a computational constraint that does not allow both Tense and Agreement features to be simultaneously

operative. Therefore they omit either tense or agreement, producing an optional sentence-final infinitive. This proposal goes beyond the case of German and Dutch: For example, it predicts that young children in English often produce forms like *him go*, using an 'object' pronoun in subject position, but only with an untensed verb. This constraint also explains other phenomena of early language development in various languages (Wexler 1998, 2003), including many phenomena that appear on the surface to be unrelated, such as the omission of object 'clitics' (i.e. weak forms of object pronouns occurring in preverbal position) in French and Italian (Jakubowicz *et al.* 1997, Jakubowicz *et al*. 1998, Hamann 2002, Paradis *et al.* 2005/2006).

5.3.   Development

One developmental hypothesis is that the computational constraint 'grows away' as a result of maturation, under genetic guidance. More traditional psychological hypotheses about learning do not seem to be adequate explanations. For example, input to the child overwhelmingly has tensed verbs in simple clauses, so children are not imitating what they mostly hear. The child's tenseless forms are not 'simpler'—they are often more superficially complex than the correct tensed forms, for example, the correct *werk* 'work' is replaced by *werken*. Standard environmental variables that are known to increase learning (Huttenlocher *et al.* 1991) have no such effect on the development of tense (Rice *et al*. 1998).

Behavioral genetic studies provide evidence for the heritability of the use of tensed constructions. A study of typically developing twins showed that the development of tense is closer in identical twins than in fraternal twins (Ganger *et al.* 1997). Further support for the biological basis for tense comes from studies of Specific Language Impairment (SLI), a condition in which children have difficulties in language, often with no apparent general cognitive deficit: They show frequent tense omission (Rice *et al.* 1995, Rice & Wexler 1996, Wexler 1996, 2003, Rice *et al*. 1998, Wexler *et al.* 2004). SLI is a genetically caused difficulty, on our interpretation involving a persistence of the early computational constraints. In fact the use of optional infinitives extends at least through the teen years, suggesting that the tense deficit may be permanent (Rice *et al.* 2009). A large study of 6-year-old twins at risk for language delay (Bishop *et al.* 2005) measured subjects' ability to repeat nonsense words, a test of the use of tense, and several other behaviors, including vocabulary size. Figure 2 shows their results.

Genetics influences on language deficits

Figure 2: Proportions of variance in language-deficit status attributable to genes (h^2_g), environmental influences shared by both twins (c^2_g) or other influences (e^2_g). The 'other' term incorporates influences traditionally referred to as nonshared environment and includes measurement error. WASI, Wechsler Abbreviated Scale of Intelligence; STM, short-term memory.
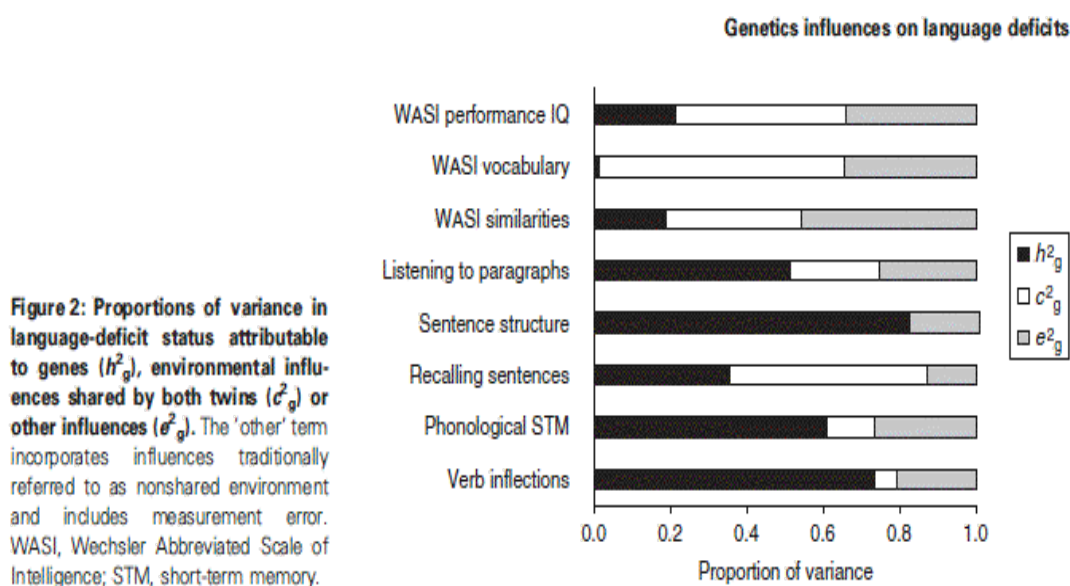
Figure 2: Proportion of variance in language-deficit status

The vocabulary measure WASI (Wechsler Abbreviated Scale of Intelligence) has almost no heritable component. Both phonological short-term memory (STM) (the ability of people to hold uninterpreted phonological strings in mind, usually measured by how well they repeat nonsense words) and frequency of the use of tense ('inflections') have very strong heritable components. But the bivariate heritability of phonological STM and tense was close to zero. Thus, the authors argue, the development of tense is independent from pure phonological memory.

There are probably two distinct forms of SLI, one related more to memory impairment, and one related more to linguistic impairment (including tense use). Several studies (The SLI Consortium 2002, 2004) have linked phonological memory to a region (*SLI1*) of chromosome 16 while Falcaro *et al.* (2008) have linked tense development to a region *SLI2* of chromosome 19. We can cautiously infer from this result that there is a chromosomal region somehow related to the optional infinitive genotype, perhaps to the growing away of the computational constraint. These methods and these kinds of data show that it is possible to find (regional) linkage for a particular aspect of language processes. Hypotheses developed in linguistic theory led to studies of the development of tense deficit in children and the discovery of the optional infinitive stage that led to the discovery of the extended nature of that stage in SLI children which led to linkage of variation in these properties to regions of the genome. Each of these steps involved creative research, but they exemplify what we hope for the future.

6.   Normal Variation in the Neurology of Language and the Genome

The study of *FOXP2* and SLI exemplifies the common method of exploring genes' potential functions by noting the disastrous impact their mutations can have on

normal behavior or structure. In the case of studying the genetics of language in humans, we want to focus on just those cases that, according to numerous publications on these pathologies, spare general cognition and reduce language (SLI, in general) or conversely reduce general cognition and spare language (e.g., Williams syndrome; see Bellugi *et al.* 1994, Bellugi *et al.* 2001, Clahsen & Temple 2003, Zukowski 2005). But a gap in one kind of human ability allows behavioral and cognitive compensatory mechanisms to come into play, thus obscuring or at least confounding the data. Consequently, it is useful to also consider cases of 'normal' variation in language representation related to genetic variation, whenever possible. This possibility has recently arisen. It involves the neurological organization for language as a function of left- vs. right-handed family background.

Language scientists have long struggled with the problem of how our internal dictionary (the lexicon) is processed in relation to syntactic composition: Is the lexicon distinct from syntactic computation? How are words integrated with syntax? Recent explorations of potential genetic differences in how words are stored shows how biology and language can fruitfully interact and further dissect the general grammatical scheme in (1) all the way to genomics. This comes from current research on language acquisition and on the brain's division of the cerebral cortex into left and right halves. A familiar biological-language asymmetry is that almost all right-handed people possess strong left hemisphere lateralization for syntactic function. However, research on certain kinds of aphasia—the pathological traumatic inability to produce or comprehend language—has revealed that right-handers with left-handed family members ('mixed families') display more right hemisphere language involvement than right-handers whose other family members are only right-handed ('pure families'; see Luria 1970, Hutton *et al.* 1977). More recent behavioral research has shown that individuals from mixed families access individual words more readily than sentence structure, while the reverse is true for pure family right-handers: Accordingly, the right hemisphere's language involvement may be specific to the lexicon (Bever 1983, Bever *et al.* 1987, Bever *et al.* 1989, Townsend *et al.* 2001). Furthermore, the critical language learning period for mixed family right-handers comes earlier than for individuals from purely right-handed families (Ross & Bever 2004), possibly because mixed family right-handers base their language acquisition learning on words as opposed to syntax. This conjecture has found recent support in an fMRI study showing greater right hemi-sphere activation for a lexical task in right-handers with familial left-handedness (see discussion in Bever 2009).

These findings on laterality confirm the basic hypothesis that mixed family right-handers have more distributed lexical knowledge. The next step will be to unravel the behavior and brain activation patterns of mixed family right-handers with and without the gene markers recently associated with left-handedness (Francks *et al.* 2007). The literature focusing on syndromes such as schizophrenia and Alzheimer's provides some models of studying normal cognitive variation in relation to genetic variation (Egan *et al.* 2003, DeYoung *et al.* 2008, Green *et al.* 2008, Tan *et al.* 2008). The case of familial handedness will be complete if only those with left-handedness genes exhibit lexically focused behavior, opening a

new window into language's genetic base and its variation in an ostensibly normal population. In addition, further research remains to demonstrate that specifically syntactic information about words is bilaterally accessed, not just word associations: For example, while *sneeze* is syntactically an intransitive verb, 'sneeze' also has strong semantic associations with 'flu' and 'sick'. Words might still be represented bilaterally in people with familial left-handedness, with syntactic information represented in the left hemisphere and associative information in the right. However it plays out, such case studies will expand our scientific understanding of the interplay between the genome, the brain, language behavior, and grammar, sharpening our understanding of the genetic endowment for language.

7.   Conclusion

Advances in theoretical linguistics, comparative ethology, genetics, and language evolution contribute to our knowledge of the biological basis of language and pave the way to what is yet to be explored in Biolinguistics. No doubt, many of the recent discoveries and theories will be modified by ongoing research, logical considerations, and new methodologies. The main point of this review is to show how the study of language can be integrated around scientific questions familiar to biologists and ethologists. We started with a formal analysis of what language *is*. These features were then discussed in light of studies of animal behavior, both natural and experimentally induced. Next, we turned to more direct genetic investigations of language dysfunctions, first in a familial phenotype, then a developmental one. Finally, we considered potential genetic influences on the neurological organization of language. Thus, we have outlined a range of typical biological methods as applied to the integrated study of language as a biological phenomenon with a critical genetic component. The results are still fragmented and subject to revision. But we are confident that a coherent picture of language as a biologically rooted phenomenon will emerge out of investigations of these kinds.

The Biolinguistic Program links language and biology in a natural way. Moreover, it proposes architectural properties of FLN and properties of the discrete infinity of human language that serve as a useful guide for further investigation of other human systems, of non-human species and their neurological organization. Additional questions arise however: How can the recursive hierarchical structures derived by Merge be further tested in humans? And what evidence can be brought about for its presence/absence in animals? While a few preliminary results are available, further neurobiological experiments are necessary. Moreover, we have pointed out some promises and pitfalls of modern genomics, and further work is needed to understand the pathways from genes to linguistic phenomena. Addressing these questions and formulating hypotheses that are testable on different populations will pave the way to a further understanding of the biology of human language.

References

Ajdukiewicz, Kazimierz. 1935. Die syntaktische Konnexität. *Studia Philosophica* 1, 1–27.

Arnold, Kate & Klaus Zuberbühler. 2006. Semantic combinations in primate calls. *Nature* 441, 303.

Avrutin, Sergey & Kenneth Wexler. 1992. Development of Principle B in Russian: Coindexation at LF and coreference. *Language Acquisition* 2, 259–306.

Bar-Hillel, Yehoshua. 1953. A quasi-arithmetical notation for syntactic description. *Language* 29, 47–58.

Belletti, Adriana & Luigi Rizzi (eds.). 2002. *Noam Chomsky, On Nature and Language*. Cambridge: Cambridge University Press.

Bellugi, Ursula, Marie St. George & Albert M. Galaburda (eds.). 2001. *Journey from Cognition to Brain to Gene: Perspectives from Williams Syndrome*. Cambridge, MA: Bradford Books/MIT Press.

Bellugi, Ursula, Paul P. Wang & Terry L. Jernigan. 1994. Williams syndrome: an unusual neuropsychological profile. In Sarah H. Broman & Jordan Grafman (eds.), *Atypical Cognitive Deficits in Developmental Disorders: Implication for Brain Function*, 23–56. Mahwah, NJ: Erlbaum.

Berglund, Jonas, Katherine S. Pollard & Matthew T. Webster. 2009. Hotspots of biased nucleotide substitutions in human genes. *Public Library of Science Biology* 7: e26.

Berwick, Robert C. & Amy S. Weinberg. 1986. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition*. Cambridge, MA: MIT Press.

Berwick, Robert C. & Noam Chomsky. In press. The biolinguistic program: The current state of its evolution and development. In Di Sciullo & Boeckx (eds.).

Berwick, Robert C. In press. *Syntax facit saltum redux*: Biolinguistics and the leap to syntax. In Di Sciullo & Boeckx (eds.).

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition and Language Development*, 277–360. New York: Wiley & Sons.

Bever, Thomas G. 1983. Cerebral lateralization, cognitive asymmetry, and human consciousness. In Ellen Perecman & Jason W. Brown (eds.), *Cognitive Processing in the Right Hemisphere*, 19–39. New York: Academic Press.

Bever, Thomas G. 2009. The individual and universal in language. In Piattelli-Palmarini, Uriagereka & Salaburu (eds.), 278–295.

Bever, Thomas G., Caroline Carrithers & David J. Townsend. 1987. A tale of two brains — Or: The sinistral quasimodularity of language. *Proceedings of the Ninth Annual Cognitive Science Society Meetings*, 764–773.

Bever, Thomas G., Caroline Carrithers, Wayne Cowart & David J. Townsend. 1989. Language processing and familial handedness. In Albert M. Galaburda (ed.), *From Neurons to Reading*, 331–357. Cambridge, MA: MIT Press.

Bishop, Dorothy V. M., Caroline V. Adams & Courtenay F. Norbury. 2005. Distinct genetic influences on grammar and phonological short-term memory deficits: Evidence from 6-year-old twins. *Genes, Brain and Behavior* 5, 158–169.

Boeckx, Cedric & Massimo Piattelli-Palmarini. 2005. Language as a natural object — Linguistics as a natural science. *The Linguistic Review* 22, 447–466.

Boeckx, Cedric. 2006. *Linguistic Minimalism: Origins, Concepts, Methods, and Aims.* New York: Oxford University Press.

Boeckx, Cedric. 2008. *Bare Syntax*. Oxford: Oxford University Press.

Bosman, Conrado, Ricardo Garcia & Francisco Aboitiz. 2004. FoxP2 and the language working-memory system. *Trends in Cognitive Sciences* 8, 251–252.

Carroll, Sean B. 2005. *Endless Forms Most Beautiful: The New Science of Evo–Devo and the Making of the Animal Kingdom*. New York: W.W. Norton and Co.

Carroll, Sean B., Jennider K. Grenier & Scott D. Weatherbee. 2001. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden, MA: Blackwell.

Cherniak, Christopher, Zekeria Mokhtarzada, Raul Rodriguez-Esteban & Kelly Changizi. 2004. Global optimization of cerebral cortex layout. *The National Academy of Sciences* 101, 1081–1086.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2, 113–23.

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

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language,* 1–52. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 2009a. Opening remarks. In Piattelli-Palmarini, Uriagereka & Salaburu (eds.), 13–43.

Chomsky, Noam. 2009b. Conclusion. In Piattelli-Palmarini, Uriagereka & Salaburu (eds.), 379–409.

Clahsen, Harald & Christine Temple. 2003. Words and rules in Williams Syndrome. In Yonata Levy & Jeannette E. Schaeffer (eds.), *Language Competence across Populations: Toward a Definition of Specific Language Impairment,* 323–352. Hillsdale, NJ: Erlbaum.

Corballis, Michael C. 2004. FoxP2 and the mirror system. *Trends in Cognitive Science* 8, 95–96.

DeYoung, Collin, G., Joseph L. Flanders & Jordan B. Peterson. 2008. Cognitive abilities involved in insight problem solving: An individual differences model. *Creativity Research Journal* 20, 278–290.

Di Sciullo, Anna Maria (ed.). 2003a. *Asymmetry in Grammar*, vol. 1 — *Syntax and Semantics*. Amsterdam: John Benjamins.

Di Sciullo, Anna Maria (ed.). 2003b. *Asymmetry in Grammar*, vol. 2 — *Phonology, Morphology, and Acquisition*. Amsterdam: John Benjamins.

Di Sciullo, Anna Maria. 2005. *Asymmetry in Morphology*. Cambridge, MA: MIT Press.

Di Sciullo, Anna Maria. 2007. A remark on natural language and natural

language processing from a biolinguistic perspective. In Hamido Fujita & Domenico M. Pisanelli (eds.), *Frontiers in Artificial Intelligence and Applications*, 126–144. Amsterdam: ISO Press.

Di Sciullo, Anna Maria. 2010. The biolinguistics network. *Biolinguistics* 4, 149–158.

Di Sciullo, Anna Maria. In press. A biolinguistic approach to variation. In Di Sciullo & Boeckx (eds.).

Di Sciullo, Anna Maria & Cedric Boeckx (eds.). In press. *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: Oxford University Press.

Di Sciullo, Anna Maria & Dana Isac. 2008a. The asymmetry of Merge. *Biolinguistics* 2, 260–290.

Di Sciullo, Anna Maria & Dana Isac. 2008b. Movement chains at the interfaces. *The Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 53, 181–219.

Duret, Laurent. 2009. Mutation Patterns in the human genome: more variable than expected. *Public Library of Science Biology* 7: e28.

Egan, Michael F., Masami Kojima, Joseoh H. Callicott , Terry E. Goldberg, Bhaskar S., Kolachana, Alessandro Bertolino, Eugene Zaitsev, Bert Gold, David Goldman, Michael Dean, Bai Lu, Daniel R. Weinberger. 2003. The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* 112, 144–145.

Enard, Wolfgang, Molly Przeworski, Simon E. Fisher, Cecilia S.L. Lai, Victor Wiebe, Takashi Kitano, Anthony P. Monaco & Svante Pääbo. 2002. Molecular evolution of FoxP2, a gene involved in speech and language. *Nature* 418, 869–872.

Endress, Ansgar D. Sarah Carden, Elisabetta Versace and Marc D. Hauser. 2009. The apes' edge: Positional learning in chimpanzees and humans. *Animal Cognition*. DOI 10.1007/s10071-009-0299-8.

Endress, Ansgar D., Donal Cahill, Stefanie Block, Jeffrey Watumull and Marc D. Hauser. 2009. Affixation in a non-human primate: Evidence of an evolutionary precursor to human language. *Biology Letters* 5, 749–751.

Engor, Roian, Cory Miller & Marc D. Hauser. 2004. Nonhuman primate communication. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 2nd edn. Amsterdam: Elsevier.

Falcaro, Milena, Andrew Pickles, Dianne F. Newbury, Laura Addis, Emma Banfield, Simon E. Fisher, Anthony P. Monaco, Zoe Simkin, Gina Conti-Ramsden & the SLI Consortium. 2008. Genetic and phenotypic effects of phonological short term memory and grammatical morphology in specific language impairment. *Genes, Brain and Behavior* 7, 393–402.

Fisher, Simon E. & Constance Scharff. 2009. FoxP2 as a molecular window into speech and language. *Trends in Genetics* 25, 166–177.

Fisher, Simon E., Faraneh Vargha-Khadem, Kate E. Watkins, Anthony P. Monaco & Marcus E. Pembrey. 1998. Localisation of a gene implicated in a severe speech and language disorder. *Nature Genetics* 18, 168–170.

Fisher, Simon E., Cecilia S.L. Lai & Anthony P. Monaco. 2003. Deciphering the genetic basis of speech and language disorders. *Annual Reviews in Neuroscience* 26, 57–80.

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

Fitch, W. Tecumseh. 2010. Three meanings of recursion: Key distinctions for biolinguistics. In Richard Larson, Viviane Deprez & Hiroko Yamakido (eds.), *The Evolution of Human Language: Biolinguistic Perspectives*, 73–90. Cambridge: Cambridge University Press.

Francks, Clyde, Shinji Maegawa, Juha Laurén, Brett S. Abrahams, Antonio Velayos-Baeza, Sarah E. Medland, Stefano Colella, Matthias Groszer, E. Z. McAuley, Tara M. Caffrey, Tõnis Timmusk, Priit Pruunsild, Indrek Koppel, Penelope A. Lind, Noriko Matsumoto-Itaba, Jérôme Nicod, Lan Xiong, Ridha Joober, Wolfgang Enard, Benjamin Krinsky, Eiji Nenba, Alex J. Richardson, Brien P. Riley, Nicholas G. Martin, Stephen M. Strittmatter, Hans-Jürgen Möller, Dan Rujescu, David St Clair, Pierandrea Muglia, J. Louw Roos, Simon E. Fisher, Richard Wade-Martins, Guy A. Rouleau, John F. Stein, Maria Karayiorgou, David H. Geshwind, Joannis Ragoussis, Kenneth S. Kendler, Matti S. Airaksinen, Mitsuo Oshimura, Lynn E. DeLisi & Anthony P. Monaco. 2007. LRRTM1 on chromosome 2p12 is a maternally suppressed gene that is associated paternally with handedness and schizophrenia. *Molecular Psychiatry* 12, 1129–1139.

French, Catherine A., Matthias Groszer, Christopher Preece, Anne-Marie Coupe, Klaus Rajewsky & Simon E. Fisher. 2007. Generation of mice with a conditional Foxp2 null allele. *Genesis* 45, 440–446.

Friederici, Angela D. 2009. The brain differentiates hierarchical and probabilistic grammars. In Piattelli-Palmarini, Uriagereka & Salaburu (eds.), 184–194.

Friederici, Angela D., Christian J. Fiebach, Matthias Schlesewsky, Ina D. Bornkessel & D. Yves von Cramon. 2006. Processing linguistic complexity and grammaticality in the left frontal cortex. *Cerebral Cortex* 16, 1709–1717.

Friedman, Robin, Farh, Kyle, Burge, Christopher, & David Bartel. 2009. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research,* 19, 92–105.

Fujita, Eriko, Yuko Tanabe, Akira Shiota, Masatsugu Ueda, Kiyotaka Suwa, Mariko Y. Momoi & Takashi Momoi. 2008. Ultrasonic vocalization impairment of Foxp2 (R552H) knockin mice related to speech-language disorder and abnormality of Purkinje cells. *Proceedings of the National Academy of Sciences of the United States of America* 105, 3117–3122.

Ganger, Jennifer, Kenneth Wexler & Melanie Soderstrom. 1997. The genetic basis for the development of tense: A preliminary report on a twin study. *BUCLD* 22, 224–234.

Gelderen, Elly van. 2004. *Grammaticalization as Economy*. Amsterdam: John Benjamins.

Gelderen, Elly van. 2008. Economy of merge and grammaticalization: Two steps in the evolution of language. In Regine Eckhardt, Gerhard Jäger & Tonjes Veenstra (eds.), *Variation, Selection, Development: Probing the Evolutionary Model of Language Change*, 179–197. Berlin: Mouton de Gruyter.

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

Geschwind, Norman & Albert M. Galaburda. 1985. Cerebral lateralization: Biological mechanisms, associations, and pathology: 11. A hypothesis and a program for research. *Archives* of *Neurology* 42, 521–522.

Givón, Talmy. 2002. *Bio-Linguistics*. Amsterdam: John Benjamins.

Gopnik, Mirna. 1990. Feature-blind grammar and dysphagia [letter]. *Nature* 344, 715.

Green, Adam E., Marcus R. Munafò, Colin G. DeYoung, John A. Fossella, Jin Fan & Jeremy R. Gray. 2008. Using genetic data in cognitive neuroscience: From growing pains to genuine insights. *Nature Reviews Neuroscience* 9, 710–720.

Grodzinksy, Yosef & Andrea Santi, 2008. The battle for Broca's region. *Trends in Cognitive Sciences* 12, 474–480.

Groszer, Matthias, David A. Keays, Robert M.J. Deacon, Joseph de Bono, Shweta Prasad-Mulcare, Simone Gaub, Muriel G. Baum, Catherine A. French, Jérôme Nicod, Julie A. Coventry, Wolfgang Enard, Martin Fray, Steve D.M. Brown, Patrick M. Nolan, Svante Pääbo, Keith M. Channon, Rui M. Costa, Jens Eilers, Günter Ehret, J. Nicholas P. Rawlins & Simon E. Fisher. 2008. Impaired motor learning and synaptic plasticity in mice carrying a point mutation implicated in human speech deficits. *Current Biology* 18, 354–362.

Guasti, Maria Teresa. 2002. *Language Acquisition: The Growth of Grammar*. Cambridge, MA: MIT Press.

Haegeman, Liliane. 1994. Root infinitives, tense and truncated structure. *Language Acquisition* 4, 205–255.

Haesler, Sebastian, Christelle Rochefort, Benjamin Georgi, Pawel Licznerski, Pavel Osten & Constance Scharff. 2007. Incomplete and inaccurate vocal imitation after knockdown of FoxP2 in songbird basal ganglia nucleus area X. *Public Library of Science Biology* 5, e321.

Haesler, Sebastian, Kazuhiro Wada, Arpenik Nshdegan, Edward E. Morrisey, Thierry Lints, Eric D. Jarvis & Constance Scharff. 2004. FoxP2 expression in avian vocal learners and non-learners. *The Journal of Neuroscience* 24, 3164–3175.

Hailman, Jack P. & Millicent S. Ficken. 1987. Combinatorial animal communication with computable syntax: Chick-a-dee calling qualifies as 'language' by structural linguistics. *Animal Behaviour* 34, 1899–1901.

Hamann, Cornelia. 2002. *From Syntax to Discourse: Pronominal Clitics, Null Subjects and Infinitives in Child Language*. Dordrecht: Kluwer.

Hauser, Marc D. & Thomas Bever. 2008. A biolinguistic agenda. *Science* 322, 1057–1059.

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

van Heijningen, Caroline, Joos Visser, Willem Zuidema & Carel ten Cate. 2009. Simple rules can explain discrimination of putative recursive syntactic structures by a songbird species. *Proceedings of the National Academy of Sciences (USA)* 106, 20538–20543.

Hinzen, Wolfram. In press. Emergence of a systemic semantics through minimal and underspecified codes. In Di Sciullo & Boeckx (eds.).

Hinzen, Wolfram. To appear. Minimalism. In Kempson, Ruth, Nicholas Asher &

Tim Fernando (eds.), *Handbook of the Philosophy of Linguistics*. Amsterdam: Elsevier.

Hodgkinson, Alan, Emmanuel Ladoukakis & Adam Eyre-Walker. 2009. Cryptic variation in the human mutation rate. *Public Library of Science Biology* 7, e1000027.

Huber, Ludwig, U. Aust, G. Michelbach, S. Ölzant, M. Loidolt & R. Nowotny. 1999. Limits of symmetry conceptualisation in pigeons. *Quarterly Journal of Experimental Psychology* 4, 351–379.

Hurst, Jane A., Michael Baraitser, Elizabeth Auger, Frank Graham & Staffan Norell. 1990. An extended family with a dominantly inherited speech disorder. *Developmental Medicine and Child Neurology* 32, 352–355.

Huttenlocher, Janellen, Wendy Haight, Anthony S. Bryk, Michael H. Seltzer & Thomas Lyons. 1991. Early vocabulary growth: Relation to language input and gender. *Developmental Psychology* 27, 236–248.

Hutton, J. Thomas, Natalia Arsenina, Bella Kotik & Alexander R. Luria. 1977. On the problems of speech compensation and fluctuating intellectual performance. *Cortex* 13, 195–207.

Jakubowicz, Celia, Lea Nash, Catherine Rigaut & Christophe-Loic Gerard. 1998. Determiners and clitic pronouns in French-speaking children with SLI. *Language Acquisition* 7, 113–160.

Jakubowicz, Celia, Natascha Müller, Beate Riemer & Catherine Rigaut. 1997. The case of subject and object omission in French and German. *BUCLD* 31, 331–342.

Jarvis, Erich D. 2004. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Science* 1016, 749–777.

Jarvis, Erich D. 2006. Selection for or against vocal learning in birds and mammals. *Ornithological Science* 5, 5–14.

Jenkins, Lyle (ed.). 2004. *Variation and Universals in Biolinguistics*. Amsterdam: Elsevier.

Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press.

Jenkins, Lyle. In press. Biolinguistic investigations: Genetics and dynamics. In Di Sciullo & Boeckx (eds.).

Jenkins, Lyle. In press. Biolinguistics: The 'third factor' in evolution and variation. In Di Sciullo & Boeckx (eds.).

Kaminski, Juliane, Josep Call & Julia Fischer. 2004. Word learning in a domestic dog: Evidence for 'Fast Mapping'. *Science* 304, 1682–1683.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Kayne, Richard S. 2002. Pronouns and their antecedents. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program*, 133–166. Malden, MA: Blackwell.

Kayne, Richard S. 2005. Some notes on comparative syntax, with special reference to English and French. In Guglielmo Cinque & Richard S. Kayne (eds.), *The Oxford Handbook of Comparative Syntax*, 3–69. New York: Oxford University Press. [Reprinted as chap. 12 in Kayne, Richard S. 2005. *Movement and Silence*. New York: Oxford University Press.]

Kayne, Richard S. In press. The limits of computation. In Di Sciullo & Boeckx

(eds.).

Lai, Cecilia S.L., Simon E. Fisher, Jane A. Hurst, Elaine R. Levy, Shirley Hodgson, Margaret Fox, Stephen Jeremiah, Susan Povey, D. Curtis Jamison, Eric D. Green, Faraneh Vargha-Khadem & Anthony P. Monaco. 2000. The SPCH1 region on human 7q31: Genomic characterization of the critical interval and localization of translocations associated with speech and language disorder. *The American Journal of Human Genetics* 67, 357–368.

Lai, Cecilia S.L., Simon E. Fisher, Jane A. Hurst, Faraneh Vargha-Khadem & Anthony P. Monaco. 2001. A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413, 519–523.

Lambek, Joachim. 1958. Mathematics of sentence structure. *The American Mathematical Monthly* 65, 154–170.

Langendoen, D. Terence. 2003. Merge. In Andrew Carnie, Heidi Harley & MaryAnn Willie (eds.), *Formal Approaches to Function in Grammar: In Honor of Eloise Jelinek*, 307–318. Amsterdam: John Benjamins.

Lasnik, Howard. 1999. *Minimalist Analysis*. Malden, MA: Blackwell.

Lasnik, Howard. 2002. The minimalist program in syntax. *Trends in Cognitive Science* 6, 432–437.

Lasnik, Howard. 2003. *Minimalist Investigations in Linguistic Theory*. London: Routledge.

Lasnik, Howard. In press a. Minimalism. In Patrick C. Hogan (ed.), *The Cambridge Encyclopedia of the Language Sciences*. Cambridge: Cambridge University of Press.

Lasnik, Howard. In press b. What kind of computing device is the human language faculty? Di Sciullo & Boeckx (eds.).

Lenneberg, Eric H. 1967. *Biological Foundations of Language*. New York: John Wiley.

Li, Gang, Jinhong Wang, Stephen J. Rossiter, Gareth Jones & Shuyi Zhang. 2007. Accelerated FoxP2 evolution in echolocating bats. *Public Library of Science One* 2, e900.

Luria, Alexander R. 1970. *Traumatic Aphasia: Its Syndromes, Psychology and Treatment* [transl. from Russian; Moscow: Academy of Medical Sciences, 1947]. The Hague: Mouton.

Martin, Roger & Juan Uriagereka. 2000. Some possible foundations of the minimalist program. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 1–30. Cambridge, MA: MIT Press.

McGilvray, James. 2005. Meaning and creativity. In James McGilvray (ed.), *The Cambridge Companion to Chomsky*, 204–222. Cambridge: Cambridge University Press.

McGilvray, James. 2006. On the innateness of language. In Robert J. Stainton (ed.), *Contemporary Debates in Cognitive Science*, 97–112. Malden, MA: Blackwell.

Medeiros, David P. 2008. Optimal growth in phrase structure. *Biolinguistics* 2, 152–195.

Miller, George A. & Noam Chomsky. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of*

Mathematical Psychology 2, 419–491. New York: Wiley.

Mody, Shilpa & Tecumseh Fitch. Submitted. Humans can learn an artificial mildly context sensitive grammar. *Cognition.*

Montell, Denise J. 2008. Morphogenetic cell movements: diversity from modular mechanical properties. *Science* 322, 1502–1505.

Moro, Andrea. 2000. *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.

Moro, Andrea. 2008. *The Boundaries of Babel: The Brain and the Enigma of Impossible Languages*. Cambridge, MA: MIT Press.

Niyogi, Partha. 2006. *The Computational Nature of Learning and Evolution*. Cambridge, MA: MIT Press.

Niyogi Partha & Robert C. Berwick. 1997. A dynamical systems model of language change. *Complex Systems* 11, 161–204.

Niyogi, Partha & Robert C. Berwick. 2009. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences of the United States of America* 106, 10124–10129.

Nowak, Martin A., Natalia L. Komarova & Partha Niyogi. 2002. Computational and evolutionary aspects of language. *Nature* 417, 611–617.

Paradis, Johanne, Martha Crago & Fred Genesee. 2005/2006. Domain-specific versus domain-general theories of the deficit in SLI: Object pronoun acquisition by French-English bilingual children. *Language Acquisition* 13/14, 33–62.

Petrides, Michael & Deepak N. Pandya. 1994. Comparative architectonic analysis of the human and the macaque frontal cortex. In François Boller & Jordan Grafman (eds.), *Handbook of Neuropsychology*, vol. 9, 17–58. Amsterdam: Elsevier.

Piattelli-Palmarini, Massimo & Juan Uriagereka. 2004. The immune syntax: The evolution of the language virus. In Lyle Jenkins (ed.), *Variations and Universals in Biolinguistics*, 341–377. Amsterdam: Elsevier North-Holland.

Piattelli-Palmarini, Massimo & Juan Uriagereka. 2005. The evolution of the narrow faculty of language: The skeptical view and a reasonable conjecture. *Lingue e Linguaggio* IV, 1–52.

Piattelli-Palmarini, Massimo & Juan Uriagereka. 2008. Still a bridge too far? Biolinguistic questions for grounding language on brains. *Physics of Life Reviews* 5, 207–224.

Piattelli-Palmarini, Massimo & Juan Uriagereka. In press. A geneticist's dream, a linguist's nightmare: The case of FoxP2. In Di Sciullo & Boeckx (eds.).

Piattelli-Palmarini, Massimo, Juan Uriagereka & Pello Salaburu (eds.). 2009. *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. Oxford: Oxford University Press.

Poeppel, David & Kenneth Wexler. 1993. The full competence hypothesis. *Language* 69, 1–33.

Raff, Rudolf A. 1996. *The Shape of Life: Genes, Development, and the Evolution of Animal Form*. Chicago: University of Chicago Press.

Rice, Mabel L. & Kenneth Wexler. 1996. Toward tense as a clinical marker of specific language impairment in English-speaking children. *Journal of Speech and Hearing Research* 39, 1239–1257.

Rice, Mabel L., Kenneth Wexler & Patricia L. Cleave. 1995. Specific language

impairment as a period of extended optional infinitives. *Journal of Speech and Hearing Research* 38, 850–863.

Rice, Mabel L., Kenneth Wexler & Scott Hershberger. 1998. Tense over time: The longitudinal course of tense acquisition in children with SLI. *Journal of Speech and Hearing Research* 41, 1412–1431.

Rice, Mabel L., Lesa Hoffman & Kenneth Wexler. 2009. Judgments of omitted *BE* and *DO* in questions as extended finiteness clinical markers of SLI to fifteen years: A study of growth and asymptote. *Journal of Speech and Hearing Research* 52, 1417–1433.

Ross, Danielle S. & Thomas G. Bever. 2004. The time course for language acquisition in biologically distinct populations: Evidence from deaf individuals. *Brain and Language* 89, 115–121.

Schroeder, Diane & Richard M. Myers, 2008. Multiple transcription start sites for FoxP2 with varying cellular specificities. *Gene* 413, 42–48.

Seyfarth, Robert M., Dorothy L. Cheney & Thore J. Bergman. 2005. Primate social cognition and the origins of language. *Trends in Cognitive Sciences* 9, 264–266.

Shu, Weiguo, Julie Y. Cho, Yuhui Jiang, Minhua Zhang, Donald Weisz, Gregory A. Elder, James Schmeidler, Rita De Gasperi, Miguel A. Gama Sosa, Donald Rabidou, Anthony C. Santucci, Daniel Perl, Edward Morrisey & Joseph D. Buxbaum. 2005. Altered ultrasonic vocalization in mice with a disruption in the Foxp2 gene. *Proceedings of the National Academy of Sciences of the United States of America* 102, 9643–9648.

Stromswold, Karin. 2001. The heritability of language: A review and meta-analysis of twin, adoption, and linkage studies. *Language* 77, 647–723.

Stromswold, Karin. 2005. Genetic specificity of linguistic heritability. In Anne Cutler (ed.), *Twenty-First Century Psycholinguistics: Four Cornerstones*, 121–140. Mahwah, NJ: Lawrence Erlbaum Associates.

Stromswold, Karin. 2006. Why aren't identical twins linguistically identical? Genetic, prenatal and postnatal factors. *Cognition* 101, 333–384.

Stromswold, Karin. 2008. The genetics of speech and language impairments. *The New England Journal of Medicine* 359, 2381–2383.

Stromswold, Karin. 2010. Genetics and the evolution of language: What genetic studies reveal about the evolution of language. In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *Evolution of Language: Biolinguistic Perspectives*, 176–193. Cambridge: Cambridge University Press.

Suzuki, Ryuji, John R. Buck & Peter L. Tyack. 2006. Information entropy of humpback whale songs. *Journal of the Acoustical Society of America* 119, 1849–1866.

Swaddle, John P., Douglas A. Ruff, Laura C. Page, Alicia M. Frame & Victoria A. Long. 2008. A test of receiver perceptual performance: European starlings' abilities to detect asymmetry in a naturalistic trait. *Animal Behaviour* 76, 487–495.

Tan, Hao-Yang, Kristin K. Nicodemus, Qiang Chen, Zhen Li, Jennifer K. Brooke, Robyn Honea, Bhaskar S. Kolachana, Richard E. Straub, Andreas Meyer-Lindenberg, Yoshitasu Sei, Venkata S. Mattay, Joseph H. Callicott & Daniel R. Weinberger. 2008. Genetic variation in AKT1 is linked to dopaminergic prefrontal cortex structure and function in humans. *Journal of Clinical*

Investigation 118: 2200–2208.

Teramitsu, Ikuko & Stephanie A. White. 2006. FoxP2 regulation during undirected singing in adult songbirds. *The Journal of Neuroscience* 26, 7390–7394.

Teramitsu, Ikuko, Lili C. Kudo, Sarah E. London, Daniel H. Geschwind & Stephanie A. White. 2004. Parallel *PoxP1* and *FoxP2* expression in songbird and human brain predicts functional interaction. *The Journal of Neuroscience* 13, 3152–3163.

Terrace, Herbert S., Laura-Ann Petitto, Richard J. Sanders & Thomas G. Bever. 1979. Can an ape create a sentence? Science 206, 891–902.

The SLI Consortium. 2002. A genome-wide scan identifies two novel loci involved in specific language impairment. *The American Journal of Human Genetics* 70, 384–398.

The SLI Consortium. 2004. Highly significant linkage to the SLI1 locus in an expanded sample of individuals affected by specific language impairment. *The American Journal of Human Genetics* 74, 1225–1238.

Thornton, Rosalind J. 1990. Adventures in Long-distance Moving: The acquisition of complex *wh*-questions. Storrs, CT: University of Connecticut dissertation.

Townsend, David J., Caroline Carrithers & Thomas G. Bever. 2001. Familial handedness and access to words, meaning and syntax during sentence comprehension. *Brain and Language* 78, 308–31.

Ullman, Michael T. & Elizabeth I. Pierpont. 2005. Specific language impairment is not specific to language: The procedural deficit hypothesis. *Cortex* 41, 399–433.

Uriagereka, Juan & Cedric Boeckx. 2007. Minimalism. In Gillian Ramchand & Charles Reiss (eds.), *The Oxford Handbook of Linguistic Interfaces*, 541–573. Oxford: Oxford University Press.

Uriagereka, Juan. 1998. *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.

Uriagereka, Juan. 1999. Multiple spell-out. In Samuel David Epstein & Norbert Hornstein (eds.), *Working Minimalism*, 251–283. Cambridge, MA: MIT Press.

Uriagereka, Juan. In press. Where to look for Fibonacci patterns in syntax. In Jordi Fortuny & Ángel Gallego (eds.), *Spelling Out Universal Grammar*. Barcelona: Universitat de Barcelona.

Vargha-Khadem, Faraneh, David G. Gadian, Andrew Copp & Mortimer Mishkin. 2005. FoxP2 and the neuroanatomy of speech and language. *Nature Reviews Neuroscience* 6, 131–138.

Vargha-Khadem, Faraneh, Kate E. Watkins, Katie J. Alcock, Paul Fletcher & Richard Passingham. 1995. Praxic and nonverbal cognitive deficits in a large family with a genetically transmitted speech and language disorder. *Proceedings of the National Academy of Sciences of the United States of America* 92, 930–933.

Vernes, Sonja C., Dianne F. Newbury, Brett S. Abrahams, Laura Winchester, Jérôme Nicod, Matthias Groszer, Maricela Alarcón, Peter L. Oliver, Kay E. Davies, Daniel H. Geschwind, Anthony P. Monaco & Simon E. Fisher. 2008. A functional genetic link between distinct developmental language disorders. *The New England Journal of Medicine* 359, 2337–2345.

Vernes, Sonja C., Jérôme Nicod, Fanny M. Elahi, Julie A. Coventry, Niamh Ken-
        ny, Anne-Marie Coupe, Louise E. Bird, Kay E. Davies & Simon E. Fisher.
        2006. Functional genetic analysis of mutations implicated in a human
        speech and language disorder. *Human Molecular Genetics* 15, 3154–3167.

Watkins Kate E., Nina F. Dronkers & Faraneh Vargha-Khadem. 2002. Behavioural
        analysis of an inherited speech and language disorder: Comparison with
        acquired aphasia. *Brain* 125, 452–464.

Webb, David M. & Jianzhi Zhang. 2005. FoxP2 in song-learning birds and vocal-
        learning mammals. *Journal of Heredity* 96, 212–216.

Wexler, Kenneth & Peter W. Culicover. 1980. *Formal Principles of Language
        Acquisition*. Cambridge, MA: MIT Press.

Wexler, Kenneth, Jeannette Schaeffer & Gerard Bol. 2004. Verbal syntax and
        morphology in Dutch normal and SLI children: How developmental data
        can play an important role in morphological theory. *Syntax* 7, 148–198.

Wexler, Kenneth. 1993. Optional infinitives, head movement and the economy of
        derivations. In David Lightfoot & Norbert Hornstein (eds.), *Verb Movement*,
        305–382. Cambridge: Cambridge University Press.

Wexler, Kenneth. 1996. The development of inflection in a biologically based
        theory of language acquisition. In Mabel L. Rice (ed.), *Toward a Genetics of
        Language*, 113–144. Mahwah, NJ: Lawrence Erlbaum.

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

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

White, Stephanie A., Simon E. Fisher, Daniel H. Geschwind, Constance Scharff &
        Timothy E. Holy. 2006. Singing mice, songbirds, and more: Models for
        FoxP2 function and dysfunction in human speech and language. *The Journal
        of Neuroscience* 26, 10376–10379.

Yang, Charles. 2002. *Knowledge and Learning in Natural Language.* Oxford: Oxford
        University Press.

Zuberbühler, Klaus. 2003. Referential signaling in non-human primates: Cogni-
        tive precursors and limitations for the evolution of language. *Advances in
        the Study of Behavior* 33, 265–307.

Zukowski, Andrea. 2005. Knowledge of constraints on compounding in children
        and adolescents with Williams syndrome. *Journal of Speech, Language, and
        Hearing Research* 48, 79–83.

Zwart, C. Jan-Wouter. 2006. Local agreement. In Cedric Boeckx (ed.), *Agreement
        Systems*, 317–339. Amsterdam: John Benjamins.

Anna Maria Di Sciullo
Université du Québec à Montréal
Department of Linguistics
P.O. Box 8888, Downtown Station
Montréal, QC H3C 3P8
Canada
Di_Sciullo.Anne-Marie@uqam.ca

Kenneth Wexler
Massachusetts Institute of Technology
Department of Brain and Cognitive Sci-
    ences & Department of Linguistics
    and Philosophy
77 Massachusetts Avenue, 46-2005
Cambridge, MA 02139
USA
wexler@mit.edu

Cedric Boeckx
ICREA & Universitat Autònoma de Barcelona
Departament de Filologia Catalana
7Edifici B, Campus UAB
08193 Bellaterra (Barcelona)
Spain
Cedric.Boeckx@uab.cat

Juan Uriagereka
University of Maryland at College Park
Department of Linguistics
1102 Marie Mount Hall
College Park, MD 20742
USA
juan@umd.edu

Lisa Lai-Shen Cheng
Leiden Universiteit
Department of General Linguistics
2300 RA Leiden
The Netherlands
L.L.Cheng@let.leidenuniv.nl

Andrew Wedel
University of Arizona
Department of Linguistics &
    Program in Neuroscience
Tuscon, Arizona 85287
USA
wedel@u.arizona.edu

Elly van Gelderen
Arizona State University
Department of English
Tempe, Arizona 85287
USA
EllyVanGelderen@asu.edu

Massimo Piattelli-Palmarini
University of Arizona
Department of Linguistics, Program in Cognitive
    Science & Department of Psychology
Tucson, AZ 85721
USA
massimo@email.arizona.edu

Robert C. Berwick
Massachusetts Institute of Technology
Department of Electrical Engineering and
    Computer Science & Department of
    Brain and Cognitive Sciences
77 Massachusetts Avenue, 46-2005
Cambridge, MA 02139
USA
berwick@csail.mit.edu

Lyle Jenkins
Biolinguistics Institute
Cambridge, MA 02139
USA
LJenkins2@comcast.net

Karin Stromswold
Rutgers University at New Brunswick
Department of Psychology &
    Center for Cognitive Science
152 Frelinghuysen Road
Piscataway, NJ 08854
USA
karin@ruccs.rutgers.edu

Heidi Harley
University of Arizona
Department of Linguistics &
    Program in Cognitive Science
Tuscon, Arizona 85721
USA
HHarley@email.arizona.edu

James McGilvray
McGill University
Department of Philosophy
Montréal, QC H3A 2T5
Canada
James.McGilvray@mcgill.ca

Thomas G. Bever
University of Arizona
Department of Linguistics &
    Department of Psychology
Tuscon, Arizona, 85721
USA
tgb@email.arizona.edu

Viable Syntax: Rethinking Minimalist Architecture

Ken Safir

Hauser *et al.* (2002) suggest that the human language faculty emerged as a genetic innovation in the form of what is called here a 'keystone factor'—a single, simple, formal mental capability that, interacting with the pre-existing faculties of hominid ancestors, caused a cascade of effects resulting in the language faculty in modern humans. They take Merge to be the keystone factor, but instead it is posited here that Merge is the pre-existing mechanism of thought made *viable* by a principle that permits relations interpretable at the interfaces to be mapped onto c-command. The simplified minimalist architecture proposed here respects the keystone factor as closely as possible, but is justified on the basis of linguistic analyses it makes available, including a relativized intervention theory applicable across Case, scope, agreement, selection and linearization, a derivation of the A/A'-distinction from Case theory, and predictions such as why *in situ wh*-interpretation is island-insensitive, but susceptible to intervention effects.

Keywords: A'-movement; Case theory; c-command; evolution; intervention; Merge; minimalism

1. Introduction

The goal of the Minimalist Program (MP) has been to reduce syntactic operations down to the simplest possible mechanism or mechanisms that are consistent with the full complexity and variation that is manifested in natural language. Insofar as this goal is achievable, it appears to open up new opportunities for addressing why it might be that the human language faculty (HLF) as we know it to function now should have emerged with relative suddenness in the evolutionary history

of humans, as some propose. The strongest hypothesis consistent with the MP is that there is only one purely syntactic mechanism that interacts with the rest of human cognition to produce HLF, which suggests that the full complexity of grammar may have arisen by virtue of a minimal change in human cognition that had cascading and transformative effects.

From this perspective, the formal goals of the MP dovetail nicely with what may be the necessary ingredient for a satisfying explanation of the sudden emergence of HLF. Put another way, one could treat the solution to the 'sudden emergence problem' as an additional boundary condition on minimalist theorizing: Any candidate for 'the simplest mechanism' that emerges from the MP-informed normal operation of linguistic science *must also* have the right properties to serve as the minimal *keystone factor* that could suddenly coordinate disparate linguistic pre-adaptations into a functioning broad HLF. The only proposal so far that is consistent with such a requirement is that of Hauser *et al.* (2002), who suggest that the Merge operation is (what I am calling) the keystone factor that achieved this fundamental reorganization, i.e. the sudden emergence of the broad HLF (see also Chomsky 2007b).

My proposal springs from the same dovetailed concerns that inform the Hauser *et al.* (2002) proposal, but I argue that the keystone factor (KF) is not Merge, which may have been part of cognition of non-*homo sapiens*. Rather, I will argue that the KF is the *Mapping Principle* that makes Merge-generated syntax *viable*, that is, interpretable by the semantic and morpho-phonological interfaces. In other words, the ability to generate recursive embedding may have predated the HLF of modern *homo sapiens*, but that HLF only emerged with the advent of the ability to interpret structures generated by Merge.

I am assuming that many faculties of mind and body, each with their own evolutionary trajectory, turned out to be useful pre-adaptations for HLF in the broad sense, but the pace of incremental changes in these pre-adaptations appears insufficient to explain the cognitive leap to HLF that appears to distinguish modern *homo sapiens* from all predecessors. This is a broad statement which I expect any number of those expert on particular anatomical and cognitive abilities to take issue with, but on the conceptual level, it is perhaps underappreciated what must be assumed if syntactic complexity is taken to be the incremental result of natural selection.

If one accepts that humans are innately prepared to learn natural languages, then it is difficult to treat the ability to learn constructions of grammar as less than general. Otherwise, one would have to argue that the ability to learn specific complex constructions, especially those not found in every language, arose because some were genetically prepared to learn them when they had to, and they passed the ability to learn the specific structure down to their offspring. Those genetically unprepared to learn the specific construction must have perished or dwindled in the population. For example, one would have to argue that exceptional Case-marking or bare infinitives, or headless relatives with matching effects arose individually in evolutionary history and were proliferated because those who had the ability to generate some of these constructions, but not others, produced more surviving offspring, even if the offspring of these successful individuals happened never to be exposed to a language with one of

these constructions.[1] Similarly, one would have to argue that those hominids that could master both Ergative/Absolutive and Nominative/Accusative case systems would have out-reproduced those that could only master one or the other, even in parts of the world that seem devoid of one or the other construction for long periods of history. This is the sort of scenario that must be accepted if HLF, as manifested in the structural configurations linguists call constructions, grew by selected genetic accretion.

Rather, it seems much more plausible that the ability to master syntactic constructions, including many that are not in the language to which one is exposed, is general, in which case humans have the capacity for knowledge about syntactic constructions that they have never been exposed to, or that their ancestors may never have been exposed to, that is, the class of possible constructions must include many that have not been specifically selected for. If we grant this much reasoning, then most of what we experience as syntactic complexity and variety must be a consequence of more general factors, not of individual constructions of grammar added step by step by natural selection. If, indeed the emergence of HLF was sudden, then the strongest assumption is that the KF consists of a *single* change in cognitive capacity that can account for the (sudden) emergence of complexity, including complexity that is not selected for. This essay is an attempt to argue for the strongest assumption.

Returning now to the Hauser *et al.* (2002) version of the strongest assumption, no single device in the history of generative grammar has ever been adequate to achieve what must be expected of the KF, and so it is no surprise that Merge is not up to it. However to see this, it is necessary to be a bit more precise about Merge. Suppose Merge is as simple as possible, that is, it is essentially Chomsky's (2004) set-Merge (see also Seely 2006).

(1)   *Merge*

If α and β are labels or the output of Merge, then Merge of α and β yields $\{\alpha,\beta\}$.

As formulated in (1), $\{\alpha,\beta\}$ can be a term in a Merge operation, for example, with γ, to create $\{\gamma, \{\alpha,\beta\}\}$. Nothing prevents a subpart of a tree from being a term in a Merge operation since any node in a tree is either a label or the output of Merge. Chomsky (2004: 110) describes cases where Merge applies to a term that is already part of a tree as 'internal Merge' (*iMerge* henceforth):

> [Narrow syntax] is based on the free operation Merge, [the Strong Minimalist Thesis] entails that Merge of α, β is unconstrained, therefore either *external* or *internal*. Under external Merge, α and β are separate objects; under internal Merge, one is part of the other, and Merge yields the property of 'displacement', which is ubiquitous in language and must be captured in some manner in any theory […]. Accordingly, displacement is not an 'imperfection' of language; its absence would be an imperfection.

---

[1]   This is not a straw man position. See Christiansen & Chater (2008: 499) who make the following assertion: "Specifically, we adopt a Construction Grammar view of language [references omitted—KS], proposing that individual constructions consisting of words or combinations thereof are among the basic units of selection".

When nodes are Merged, neither of the nodes is changed as a result of Merge (the No Tampering Condition), thus the result of iMerge applying to a sub-constituent B of dominating X will be a copy of B in the position where B occurs before Merge. On these assumptions, as Chomsky points out, a system that did not permit iMerge would require a stipulation to prevent it from occurring (an 'imperfection'), since it comes for free with Merge as defined in (1). Let us assume for the time being that *Extension*, introduced in Chomsky (1995), is a condition on Merge.

(2)   *Extension*

        Always Merge to an undominated node.

I assume that non-terminals formed by Merge or individual morphemes bearing labels may be thought of as terms, and hence nodes, in syntax.

The formulation of Merge in (1) conditioned by Extension in (2) is very elegant, but it is not often what is assumed *in practice* in most current minimalist architectures. For example, the reduction of Move to an instance of Merge, has been obscured because residues of earlier ideas have not been reevaluated. In much current practice, iMerge and eMerge (Merge of α and β when α is not contained within β) are still distinguished by special features and triggers that formerly played the role of making more expensive iMerge possible, and by different outputs of Merge, such as pair-Merge (see Chomsky 2004: 117–122) as opposed to set-Merge (where only the latter is expressed in (1)). In many minimalist accounts, the architecture is enriched beyond Merge with Agree, *uninterpretable features, projection, numerations, percolation, Spec-Head feature-checking*, and certain combinations of these which amount to operational triggers distinguishing instances of iMerge from eMerge. If Merge is the KF, then every one of these additional linguistically specific operations or entities represent departures from KF reasoning.

Once the economic distinction between iMerge and eMerge is discarded, as it is in Chomsky (2004), and some common accretions to Merge have been set aside, only Merge and a slight revision of Extension are required to generate syntactic structure. I argue further, however, that the KF must be something other than Merge. As a result, the architecture of derivations will be quite different.

1.1.   *A New Direction*

Suppose that HLF does not consist of a separate component of human cognition, but in a change of the interface relations that occurred between pre-HLF faculties (a possibility suggested in Hauser *et al.* 2002: 1573). If a single factor is to recruit the pre-adaptations that constitute HLF broad, then it must at minimum introduce or presuppose that the pre- adaptive domains become interpenetrable by virtue of a factor that permits a common interface for (at least some) pre-existing human cognitive capacities. Suppose that Merge already existed in human cognition, but was perhaps encapsulated in another cognitive component, or formally available but dormant. Accordingly, I propose that the KF was not Merge, but a change that has the effect of making Merge *viable* by permitting

varieties of cognition to be mapped onto a common medium, the forms gener-
ated by Merge.

The essential proposal of this article, roughly put, is that *the KF is the ability
to map interface relations onto c-command relations defined on the output of Merge*.
Since c-command will be argued to be, on the one hand, independently (empiri-
cally) necessary,[2] and on the other, inexpressible unless it applies to the output of
Merge, I will proceed to explore the theoretical intuition that Merge is a largely
non-viable pre-*homo sapiens* (pre-HS) capacity, and that only the introduction of a
c-command-like notion makes complex syntax available for the integration and
expression of cognitively significant relations. The goal of this line of analysis,
respecting the burden of the KF, is to derive every aspect of syntactic architecture
relevant to HLF from the factors listed in (3).

(3)   a.   Merge;

b.   interpretatively significant relations mapped onto c-command (the
KF);

c.   lexical properties; and

d.   a preference for relative proximity that may not be specifically
linguistic.[3]

As a working hypothesis, then, any minimalist architecture that does not derive
from (3a–d) must be jettisoned in favor of architecture that respects the burden of
the KF.

A reason why c-command is an especially good candidate for the factor
that produces a cascade of complexity is that it can be thought of as a structural
way of defining *potential closeness* relations (by providing a downward vector)
between nodes that may be (potentially) indefinitely far away. C-command also
serves to define *relative* closeness (e.g., if X c-commands Y and Y c-commands Z
then Y is more local to Z than X). Local domains may then be seen as emerging
from the interaction of pre-existing conceptual relations that become expressible
in viable syntax and that are then susceptible to proximity interventions. In other
words, the existence of syntactic locality domains may be an emergent con-
sequence of syntax made viable by the KF: Mapping of interpretively significant
relations onto the potential closeness relation.[4]

---

2   In Appendix A to this article, proposals made in the literature purporting to derive the
effects of c-command from Agree are shown to be untenable, unless Agree is so expanded as
to reduce to the essence of the Mapping Principle. Even if Agree were sufficient to replace c-
command mapping as the KF, Agree is also an accretion to the theory after Merge, and
would face the same KF burden that c-command does.

3   Conditions on 'computational complexity' play a similar role of inducing locality for
Chomsky's version of minimalism, where it is also hypothesized that the relevant constraint
is not linguistically specific.

4   Of course, one can ask why c-command should be the relation to define this relationship
rather than, say, dominance. See Neeleman & van de Koot (2002) for an attempt to derive c-
command from dominance and projection of labeling. Viable syntax architecture does not
permit projection of labeling. If Neeleman & van de Koot's theory were disciplined accor-
ding to the burden of the keystone factor, assuming labeling but not c-command, an alter-
native approach to some of the argumentation for c-command presented here might be
possible, but a very different architecture would emerge, permitting relations in addition to

To grasp the substance of this proposal, as opposed to other imaginable ones, it is important to understand that viability is only partial in HLF. The structures generated by Merge can be quite complex, consisting of a great number of nodes that define subunits, depending on how large a sentence is. Syntacticians differ markedly on what exact structure they would assign to a sentence like (4), but the (node dominating the) word *him* and the (node dominating the) word *he* are not normally regarded to be in a direct relationship that is regulated by syntax, even if *he* and *him* in (4) are taken to pick out the identical object of thought.

(4)   A woman who knew *him* thought that *he* was brave.

Yet there is a structural relation between *him* and *he* that one could calculate in terms of the number of nodes that each is dominated by up to the node that dominates them both. If such a relation were viable, semantic generalizations could be stated on such a correspondence. No such relationship is viable in UG, nor are any number of other such relationships that could be defined on structures generated by Merge. It is big news in linguistic theory when new structural relations are posited to be viable. Selection under sisterhood and c-command were among the first of these, to be followed by various locality relations, not all of them still assumed to be part of modern theory, including subjacency, binding domains, maximal projections, government, and phases, to name a few. The theory developed here attempts both to limit viable relations and to permit just a small set of local relations that arise as interventions determined by relative closeness in contexts where potential closeness holds.

Suppose, then, for the sake of argument, that Merge is a pre-HS capability (which may be viable for some other cognitive domain, such as predation planning or kinship calculation), but the ability to map relations in other cognitive domains to relations between nodes is not. As suggested above, a structural relation is *viable* if and only if an interface relation is permitted to exploit it. Suppose further that 'sister' and 'dominate' are not viable relations, that is, 'X dominates A' and 'X dominates B' are not interpretively visible on their own, except for the calculation of c-command, as stated in (5).[5]

(5)   *C-Command*

       B is *c-commanded* by A if B is dominated by a sister of A.

If the structural domination relation is not viable, then no interpretive relation can map onto it, and if so, then inheritance of a label, if posited as a contentful relation spreading from A or B to dominating X, would be invisible to interpre-

---

those allowed by unique appeal to c-command. On grounds of parsimony, a c-command primitive may be preferable to one based on labeling, but this depends on how the primitives are formulated.

[5]   Seely (2006: 197) assumes that a node without a label would be uninterpretable at LF and therefore cannot exist, that is, non-terminal nodes cannot exist, hence a set-theoretic notion of containment must be preferred to dominance. However it is assumed here that LF only evaluates interpretable properties for well-formedness, not uninterpretable ones. See fn. 10.

tation. Now suppose that Merge creates [Y [$_X$ A, B]] and we call the whole structure Z, as in (6).[6]

(6)   [$_Z$ Y [$_X$ A, B]]

Neither A nor B of the composite [$_X$ A, B] would be interpretable in relation to Y, unless c-command is a viable relation. The structure in (6) is generable, but pointlessly so, because it cannot be used to shape, for example, semantic interpretation. Only the availability of the *Mapping Principle* in (7) permits interpretation to exploit relations between Y and sub-constituents A and B embedded in X, where X is the sister to Y as illustrated above in (6).

(7)   *The Mapping Principle*
        If x and y are in relation R, then R is mapped onto syntax if and only if either x c-commands y or y c-commands x.

Thus Merge may have been in the cognitive arsenal of pre-HS hominids—and if so, they had the potential to generate complex trees, possibly useful for other purposes, but could not use them for semantic or phonological interpretation until the advent of The Mapping Principle.

C-command is a powerful template for interpretive relations. For example, if A or B is a non-terminal in (6), then Y also c-commands all daughters of A and B, so the recursive structures always generable by Merge are now viable and may be exploited by interpretive relations. C-command will permit a head to be in a selection relation to something inside its sister, it will permit antecedents to antecede anaphors and a copy left by internal merge to be related to a copy that is c-commanded by it. Discourses may then be thought of as concatenations of the final outputs of Merge operations. If A and B are concatenated nodes in a discourse, then neither A nor any node inside it c-commands B nor vice versa. If the Mapping Principle as stated is the KF, then antecedent relations that are not in a c-command relation are uninterpretable by any system or component that has access only to syntactic structure.

The burden of the KF is now carried by the Mapping Principle as it makes structures created by Merge interpretable and creates a common vehicle for the integration of relations expressed by interface components. Thus all of the complexity of grammar should now be expected to follow from the formulation of Merge, interpretive relations mapped onto c-command (as the means of expressing proximity of interpretive relations), *and* the general assumption that interpretive relations can be sensitive to relative proximity, that is, to interventions.

---

[6]   Epstein *et al.* (1998) point out that if Extension ensures that each new constituent is a sister to the whole structure, then c-command relations could simply be recorded as moments in the history of a derivation (a point further developed in Seely 2006). This assumption is compatible with viable syntax architecture and I will be exploited here with respect to 'Firstness', but whether c-command is derivationally defined or not does not explain why c-command is a viable relation.

(8)   *Intervention*

       B *intervenes* between A and C if A c-commands B and C, and B c-commands
       C.

(9)   *Intervener*

       B is an *intervener* between A and C if
       i.   B intervenes between A and C, and
       ii.   for some relation *R*, *R* cannot relate A and C when B is present.

This account of local relations as bounded by intervention is essentially a restate-
ment of relativized minimality defined on c-command, as in Rizzi (1990) (not just
relative proximity of any sort, as in Koster 1976). On this account, however,
*locality only emerges when the Mapping Principle makes viable the relations on which
interventions can be defined*. Put another way, potential proximity in the form of c-
command is logically prior to locality, just as Merge is logically prior to c-
command.[7] Thus interventions do not block Merge operations, but they may
block one interpretive relationship or another from being established on the
output of Merge operations.
       Consider, for example, how intervention works if *c-selection* is relation *R*.
Since labels on nodes are only introduced into a derivation on the lexical items
(heads) that bear them (since projection relations are not viable), the complement
relation in viable syntax architecture (VSA) can only be formulated as a relation
between a selecting head and a head bearing a label that it selects. A head H
selects a head Y if H must seek Y, H c-commands Y, and there are no intervening
selectors between H and Y. So if V selects D (e.g., the verb *kill* selects for a nomi-
nal), then V must c-command D and no other selector can intervene between V
and D. Thus, *depend* could not select for *the* in *Don depended on the boat* because *on*
is a selecting head that intervenes between *depend* and *the*. The reasoning here is
in the spirit of Collins (2002), though I do not follow his proposal in detail. A
head cannot select for a non-terminal Y as its complement (e.g., a verb cannot
select for a PP) because projection is not viable, and, more specifically to VSA, H
cannot select a sister to H because sisterhood is not c-command (i.e. there are no
viable head-head sister relations). Thus a simple sentence like *John must leave*
would require that if *must* selects *leave*, then *leave* is not a sister to *must*, but is
embedded in a branching sister to *must*, as it is in all theories that assume 'little
v'.[8]

---

[7]   Frank & Shanker (2001) propose that c-command, as opposed to dominance, should be
       treated as a primitive, which could conceivably be regarded as an impetus to remove the
       logically prior status of Merge for c-command (i.e. if trees are created by the formation of c-
       command relations). As a justification for this move, they point out, as I have here, that
       many sorts of relations stated in terms of dominance and precedence might be treated as
       significant, but only c-command seems to play a significant role in syntax. Their axiom-
       atization of syntactic tree structure in terms of c-command starts from the assumption that
       sisters are in a mutually c-commanding relation, which is not assumed here. It is unclear
       what the consequences for their account would be if sisters are not in a c-command relation
       (it is also assumed in the text that domination is not reflexive, or sisters would c-command
       each other, given (5)).
[8]   Since c-command is always asymmetric in this system, no viable relation results when

Thus VSA replaces the boundary-forming function of projection with boundaries created by intervention. This approach to intervention begs what might be called the Natural Intervention Hypothesis, as stated in (10).

(10) *Natural Intervention Hypothesis*

a.   Strong Form: If B intervenes between A and C for relation *R*, then B must be eligible to participate in *R* with A and C.

b.   Weak Form: If B intervenes between A and C for relation *R*, then B must have properties in common with A and C.

Although they have not always been distinguished, both versions of the Natural Intervention Hypothesis have been explored in the literature on interventions of various kinds (e.g., Koster 1976) and even much earlier in phonology. Although (10a) makes the strongest predictions about possible interveners, it has well-known problems that will not be solved here. My proposals are consistent with at least (10b). Moreover, as I develop VSA, interventions will more typically arise from intervening heads rather than intervening 'specifiers', even though the apparent interpretive relation may be one between non-terminal nodes.[9]

The foregoing introduction is meant to introduce the following propositions:

(11) a.   The sudden emergence problem, with its subparts (single factor change, unselected complexity, recruitment of existing capacities) requires positing a KF.

b.   The KF is the advent of the Mapping Principle.

c.   Relations mapped onto structure are sensitive to a pre-linguistic preference for relatively closer relations (once c-command provides the vector).

d.   All the architecture of syntax and the restrictions that shape constructions of grammar derive from the interaction of Merge, interpretive relations stated uniquely on c-command, relative proximity, and the distinct properties of lexical items.

(11a–d) are empirical hypotheses, but (11a, c) are assumed here without argument, as they are in Hauser *et al.* (2002). My main goal is to show that an architecture for syntax consistent with (11b) is defensible, namely, the one in (11d). In defending (11d), appeal will be made to pre-existing cognitive capacities from which mappable relations arise, and those attributions are empirical hypotheses that will not be explored. An honest appraisal of other syntactic theories would reveal similarly rich assumptions about human capacities that are not defended,

---

terminals are merged. Only merger where one of the terms is non-terminal results in a c-command relation. In theories with mutual c-command (sisters), there would be unintended intervention effects (as Seely 2006: 202 points out for the theory of selection in Collins 2002).

[9]   The viable syntax instantiation of relativized minimality is more restrictive than the original formulation, insofar as no minimality restriction can be placed on intervening maximal projections, for example, node-naming locality statements like subjacency list nodes that are in the class of interveners. See section 6 on cyclicity.

but the point here is not that my assumptions should escape scrutiny. Rather, my assumptions should be put into perspective with respect to the task at hand, which is defending (11d).

1.2.   *The Road Ahead*

The program throughout is to show that VSA, as a particular instantiation of the MP, can be constructed and that it comes close to respecting (11d) and which dispenses with most of the ancillary mechanisms surrounding Merge, or else reduces those mechanisms to interpretive relations permitted by the Mapping Principle. Where the program fails to truly reduce a problem, as it occasionally does, I will try to show that VSA fares no worse than other existing minimalist accounts, but it is to be acknowledged that in the larger picture this is not good enough to meet the burden of the KF, and is just the best that I was able to do. In spite of this caveat, I will argue that VSA offers certain advantages over competing analyses based on assumptions common to most minimalist accounts. These include a relativized intervention theory applicable across Case, scope, agreement, selection and linearization, a revision of Extension consistent with late attachment and head-to-head movement, derivation of the A/A'-distinction from Case theory, and a derivation of why *in situ* wh-interpretation is island insensitive, yet susceptible to intervention effects. Whether or not one accepts the evolutionary reasoning that has inspired my proposal, my central contention is that VSA deserves to stand on its own merits as a parsimonious and insightful account of the relation of syntax to interpretation at the interfaces.

The remainder of this essay fleshes out a version of VSA that respects the KF as closely as possible. Section 2 clears the field of accretions to Merge in minimalist syntax that cannot be countenanced in VSA. Section 3 develops some major design features of VSA, further specifying of the principles for assigning interpretation to structure and the interaction between structure and interpretation in the course of a derivation, while keeping in mind new devices should not complicate the KF, or ascribe implausible properties to other components. Subsequent sections extend VSA design to Case-marking (section 4), to the derivation of the A/A'-distinction from Case Theory and the interpretive role of criterial positions (section 5), to cyclic linearization and the locality of extraction (section 6), and to contrasts between terminal and non-terminal node realization that derives the role of pied-piping on intervention effects (7). Section 8 concludes by briefly summarizing the main arguments for VSA.

2.   **What We Must Do Without**

The Hauser *et al.* (2002) proposal that Merge is the KF does not succeed because it relies, in practice, on many additional devices to achieve a descriptively adequate account of phenomena considered central to natural language grammar. The list that follows includes a range of devices, not all of them found in every minimalist proposal, but many of them found in most minimalist proposals, that do not stand up to the logic of the KF if that factor is just Merge. Residues of

earlier accounts of iMerge as distinct from Merge must also be swept away, in particular, the notion that only iMerge must be operationally triggered (e.g., by feature-checking). In presenting a list of what we must do without, I clarify what VSA must achieve.

Consider first projection. A labeling relation between a node and a head it dominates is something that can be added to the Merge operation, but it is not a consequence of simply combining two terms. Projecting a label is not a viable relation in VSA since it is not a c-command relation. All that VSA can countenance is a label on a terminal node (head) and nodes that, because they are not terminal, have no label.

If 'the feature composition of a node" has no meaning for a non-terminal in a theory without projection, then there is no way to express Spec-Head checking, which is another accretion on Merge. There is no label with features on a non-terminal node for a head to check. Since heads do not c-command their 'Spec', no viable relation holds between head to Spec either. Even if a theory with Agree does not require Spec-Head relations (e.g., Chomsky 2004), VSA goes further by rendering Spec-Head relations ineffable.[10]

Removing Spec-Head checking removes a mechanical necessity for percolation, which has been used to account for how a node dominating K can be attracted because K is in it. Percolation permits a property of a maximal projection A (containing K) to inherit feature(s) of K, and then, potentially, for the feature of A to be inherited by B, a maximal projection that contains A, and so on. Percolation is not viable in any case, since it involves mapping interpretive properties onto dominance relations and not c-command relations.[11]

---

[10]   It is usually assumed in minimalist accounts that ill-formedness in phonology or semantics might result because *an interpretable feature has been left uninterpreted*. Although I don't express this claim in terms of features, the idea that formal properties of interpretation (morpho-phonological or semantic) condition possible outputs is also a crucial part of VSA.

However, the notion 'uninterpretable feature' goes beyond these considerations and requires a retreat from the logic of the KF. It is not at all clear that any phonological feature, whether it is checked or not, should ever be visible to semantic interpretation by its very nature. Similarly, no semantic feature should necessarily be visible to phonology if unchecked. Being visible in the 'wrong' component is not a necessary assumption, and if not, then phonological features do not supply any information to the semantic component, *including information about ill-formedness*, nor can semantic features contribute to the ill-formedness of phonological representations. In other words, segregating features by component restricts their descriptive power (and thereby derives the proposal of Pesetsky & Torrego 2001, that every proposed feature should be interpretable in *some* component). Structural case features play a role in phonology, for example, but are not relevant to semantics, whereas inherent Case features, which have both morphological and semantic value, may be visible in both components. Why Case features should have an origin in some pre-linguistic component remains mysterious (see Bobaljik & Wurmbrand 2008 for a summary of the issues), and it is in the latter sense that they represent a challenge to KF reasoning.

[11]   Percolation has been stipulated to apply whenever a non-terminal of arbitrary size is moved (e.g., containing *wh*-REL, as in *Bill, Al's pictures of whom, we will soon see*), but then percolation is just a notion that describes what is empirically possible—it is not an explanation (as pointed out by Heck 2004, 2007). Watanabe (2006) actually introduces a pied-piper feature that can be probed for by Agree in order to trigger movement. VSA countenances neither the role of Agree in Merge or the feature on the targeted maximal projection. Restrictions on pied-piping are poorly understood, but one limitation may be that the *wh*-REL must covertly move to a scopal position, and cannot do so if conditions on movement are violated (see e.g.

The elimination of feature-checking and percolation from the theory on the assumption that Merge is free as long as the output is interpretable at the interfaces obviates any appeal for operational triggers embedded in the application of Merge. Now consider the function of Agree, which has been appealed to play two roles in recent years (e.g., Chomsky 2000, 2001). One is to value features of the goal (and is still widely employed for this purpose), and the other, in some versions of minimalism, is to make the goal susceptible to iMerge (to 'activate' it, in some theories, for example, Bošković 2007) in combination with an EPP feature, that is, to trigger movement (e.g., in Pesetsky & Torrego 2001). On this point Chomsky has frequently been explicit, even as late as Chomsky (2004), where he first proposed that iMerge and eMerge are the same operation.

> If there is no Spec-Head relation, then the EPP-feature OCC cannot be satisfied by Merge alone. It follows that internal Merge requires Agree. Therefore Move = Agree+pied-piping+Merge.   (Chomsky 2004: 114)

Agree, however, cannot relate a head and a maximal projection in the VSA version because there are no maximal projections in VSA. Thus, Agree is a viable relation in VSA just in case it is mapped onto a c-command relationship between terminal nodes, and I will appeal to just such a relation more than once, but it cannot be used in VSA to activate non-terminals for movement. If Chomsky (2004) is right that Merge is not constrained by economy, and if Agree and pied-piping are not added to Merge, then iMerge and eMerge should be equally possible at any point in a derivation. The result may be uninterpretable, however, and so the descriptive and explanatory burden must be borne by the architecture of syntax-sensitive interpretation.

Most minimalist accounts posit that, in addition to Merge, every derivation begins with the selection of a numeration, a fixed set of selections from the lexicon to be used in the course of a derivation and to be depleted until the set is empty. Appealing to the numeration, one could still maintain that iMerge is less economical that eMerge just in case an operation that reduces the numeration is more economical than one that does not (see e.g. Lasnik & Uriagereka with Boeckx 2005: 166 and Safir 2008: 331). However, the only reason to stipulate a numeration as part of a derivation is to insure that the most efficient derivation is computed. The actual selection of numerations is unprincipled. If interpretation, rather than economy, is the only arbiter of the well-formedness of a derivation, then pre-derivational numerations are superfluous.[12]

Sauerland & Heck 2003 for intervention effects within pied-piped constituents). This suggests that displacement is not restricted by conditions on Merge, but the result must be interpretable (a hypothesis that is fleshed out for VSA in sections 5 and 7).

[12]   In rejecting economy calculations on forms that are independent of their semantic import, I am also rejecting approaches to syntax such as that of Optimality Theory, which begin from the assumption that GEN takes *a given input of forms* and generates all the representations that are competitors for the optimal derivation or representation. I do assume that the output of Merge submitted to the interpretive component contains a set of lexical items which are mapped onto the relevant prominence relations. The 'post-derivational numeration' and the structure built on it could be used to compute optimal form-interpretation matches, as in Safir (2004a), where substitutions into the post-derivational numeration can compete with the actual output (in the spirit of Reinhart's 2006 'reference set computation').

The elimination of projection, percolation, operational triggers distinguishing eMerge and iMerge, Agree as part of Merge, Spec-Head feature-checking, and numerations clears away a lot of what compromises Merge as a model for the KF, but leaves Merge insufficiently expressive. We must ask now whether the Mapping Principle meets the empirical burdens these accretions to Merge were supposed to address, while still respecting the burden and the logic of the KF.

3.   Major Design Features of Viable Syntax Architecture

There are at least two properties of syntactic structure that do not appear to be required by a compositional semantics defined to interpret recursive structures. These include geometrically-defined locality restrictions, some of which appear absolute (nothing outside of the local domain can be accessed for interpretation), and in other cases, the distance between two potentially related nodes A and B can be unbounded, but any node of the wrong sort intervening between them will be enough to prevent A from being related to B. Both of these phenomena will be treated as forms of intervention effects in VSA.

Moreover, every theory of syntax must be able to resolve conflicts that arise between the mapping of lexical argument structure onto syntax and the mapping of scopal relations onto syntax. A typical conflict of this sort arises when argument A of predicate P in the argument structure of P(A, B) must be more prominent than argument B, but B must have scope over A for some quantificational relation. The usual syntactic approach to this conflict is to assume that argument structure projects onto prominence relations in syntax and then movement or its analog reorders arguments or their parts to achieve scopal prominence. The resolution of these conflicts may differ cross-linguistically. For example, the c-command relations that must exist to support a given word order may not appear consonant with those that must express scopal relations, as in the case of *in situ* phenomena. Any syntactic theory that aspires to adequacy must be able to both characterize and, hopefully, predict, the class of possible prominence conflicts and the class of possible resolutions. The burden of VSA is to insure that all such prominence relations can be best expressed as interpretative relations mapped onto c-command relations.

I am assuming that the prominence relations just mentioned are probably *not* part of what is introduced by the KF, but rather that the expression of them in structure is what is new. Whatever determines that agents are more prominent than patients for a given lexical argument structure, or for such argument structures generally, is not assumed here to be part of what Merge or c-command contributes. The assumption that lexical argument prominence relations for particular lexical predicates predates the advent of c-command is a strong claim, not defended here, that could easily be false. If prominence relations *only emerged* with HLF, then there is a greater burden to show that notional prominence could only emerge as a result of the KF. For example, it is not obvious that scopal

---

However, no form–interpretation competition can be part of GEN unless a great deal of semantics is built into it. See Safir (2004a: 234-237).

relations, especially relative scopal relations, are effable without sufficient syntax, since they are not prominence relations directly associated with lexical items. In other words, it is possible that relative scope could be an emergent property in VSA, but argument prominence for verbs or perhaps some asyntactic notion of domain associated with particular quantifiers may predate the KF.

3.1.   The Place of Interpretation

From the VSA perspective, Merge must provide the structures from which prominence relations can be read by the syntax sensitive interpretive component. VSA is thus in the tradition of theories that generate syntax freely and filter the output, where in this theory, the phonological and interpretive components do all of the filtering. Insofar as setting of argument prominence relations will be relative to semantic classes of heads, argument prominence will be distinct from scopal prominence, but apart from the classes of heads involved, the formal prominence-setting mechanisms will be the same for both. If argument prominence is thought of, in GB-traditional terms, as 'A-relations' and scopal prominence as 'A'-relations', then the theory of A/A'-distinction will reduce to a difference between classes of heads, but will not reside in the way prominence domains are set or trees are generated, nor will there be distinct versions of movement or adjunction that require elaboration of Merge (such as pair-Merge, as in Chomsky 2004, or probe-contingent iMerge, as in Chomsky 2007a, cited above).

For example, if we assume that for every class of quantifier, there is a scopal position (or one of a set of scopal positions) to which it must move to receive an interpretation, as in the theory of Beghelli & Stowell (1997), then a quantifier without a compatible scope domain will fail to have a proper interpretation (see section 7.1).[13] I also assume that for every predicate P with more than one argument, there is a prominence ordering between the arguments of P (e.g., *killer* and *killee* for *kill*) interpreted from syntax as a c-command relationship between those arguments. If a verb cannot recover prominence relations from the tree that match the prominence relations in its lexical entry, the verb cannot have a proper interpretation. In this way, semantic conditions on scope and argument structure filter out uninterpretable trees from amongst those that can be generated by Merge applying freely.

Another design feature that is adapted here from existing accounts is that there is only one syntactic movement component: quantifier movements must take place amongst the movements that result in overt reordering as part of the

---

[13]   The approach pursued here shares much with that of Bošković (2007) insofar as an insufficiency on a constituent is frequently what drives that constituent to move, rather than a trigger that may not have yet entered the derivation, particularly in cases of successive cyclic movement. Bošković's mechanism is different, in that he explicitly assumes that the insufficiency is an unvalued feature which Last Resort permits to move in the interest of convergence, but Last Resort is a form of look-ahead, which Bošković is trying to avoid. Still, this use of Last Resort permits Bošković to dispense with appeal to intermediate triggers for successive cyclic movement. In VSA the movement is optional, but a moved constituent must find the right domain by the time it reaches interpretation, and must avoid local interpretation by escaping to the intermediate zone (see section 6), so intermediate triggers, and triggers in general, are also unnecessary (and unstateable) in VSA.

same process. The phonological component will determine which copies generated by movement are to be pronounced. In such theories, notably that of Bobaljik (2002) and references cited there, most, if not all of the movements that are observed in any language are found in all languages, but in those languages that appear not to have a movement found in some other language, the phonology masks the presence of movement by pronouncing the lowest copy. From this perspective, wh-*in-situ* phenomena are just instances where the lowest copy has been pronounced instead of the highest one. I will not adopt these assumptions whole, as my discussion of pied-piping and intervention will show, but my design follows the same leading idea.

I also draw some of VSA design from a theory that has been regarded, in part, as a competitor to the approach just described. I assume that prominence interpretation proceeds in tandem with the formation of structure to *set prominence* and *assign domains* and that it does so in ways that later operations in a derivation cannot revise. It is convenient to give this general design hypothesis a name, as in (12).

(12) *Derivational Drag Hypothesis*

Certain relations, once established at a given point in a derivation, are never revised.

In an architecture with this design, every operation that results in, or contributes to, prominence-fixing for interpretation or for phonology limits the possible interpretations of structure introduced subsequently. In this respect, my approach adapts ways of thinking introduced by Chomsky (2000) as 'cyclic Spell-Out' and adapted in Fox & Pesetsky (2005), where precedence relations introduced in the course of a derivation constrain possible linearizations that could result from further derivational steps. My approach will also use cyclic Spell-Out to (indirectly) freeze overt displacement across more than one cycle, but also to distinguish the cyclic Spell-Out of terminals vs. non-terminals. Derivational drag effects are also posited to arise from the assignment of unerasable Case and unerasable scope. Essentially, once a node is assigned a non-Nominative structural Case, no occurrence of that node can receive any other Case assignment without inducing Case conflict that causes the derivation to crash. Only a node assigned Nominative permits subsequent copies of that node to be reassigned some other Case. If Case is assigned cyclically, only Nominative nominals can move. This system is developed in section 4 and illustrated with several derivations.

3.2.   Setting and Assigning a Domain

Derivational Drag insures the preservation of interpretive relations once they are introduced, but another key notion that resonates throughout the approach developed here is that many interpretive relations are established at the first point in the derivation where that interpretive relation can be interpreted. This means that there will be a special importance to points in a derivation where a new c-command relation is established that permits mapping of an interpretive

relation onto it. Principles for setting and assigning domains are presented in (13).

(13) a.   *Setting a Domain*

    If H is terminal, every node H c-commands when it is (first) merged is in the domain H_D.

      b.   *Assigning a Domain*

    The domain H_D of head H is assigned to the first X merged such that X c-commands H_D and H.[14]

On this analysis, domain assignment is thus a two-step process for phrasal nodes. First a head H *sets* the domain it c-commands (H_D) at the first point where H is merged. The first 'X' that c-commands H_D and H has the domain of that head *assigned* to it. Scope is assigned, for example, when a scope-marking head H sets a scopal domain H_D, and then a QP is merged to [H H_D] (where H_D is shorthand for the set of nodes c-commanded by H, that is, H_D is not the name of the node that is a sister to H).

  Although there may be reasons to doubt the parallel later, an instance of Merge that induces domain assignment may be thought of as the derivational equivalent of the specifier relation. However in this account, the uniqueness of 'specifiers' is a derived consequence of the role of the *first* c-commander of a domain, which presents the first opportunity in a derivation to map an interpretive relation, namely, domain assignment. Firstness could be factored out

---

14   Note that (13b) does not stipulate that the setting of a domain is always an asymmetric relation, although it is possible that it always is. Domain-setting is asymmetric because a terminal and a non-terminal are merged, but as it stands, (13b) allows that merger of a non-terminal with another non-terminal that could result in both constituents being assigned a domain by first c-commanding a domain in its counterpart, e.g., where merger of [X, X_D] and [Y, Y_D] would allow that Y_D is assigned X_D and X_D is simultaneously assigned Y_D. To (13b) could be added the following condition: No more than one domain is assigned on any given instance of Merge.

  As Michal Starke (p.c.) points out, there is otherwise too much symmetry for certain instances of Merge, yielding uncertainty about what sort of constituent results from the merger. I am not convinced that symmetry in such cases could not be ruled out on independent interpretive grounds, but if not, then this stipulation is necessary and recovers part of the information formerly expressed in terms of projection relations when non-terminals merge. When non-terminals merge and no domain is assigned (e.g., neither non-terminal immediately dominates a domain set by a head), it is not clear that any asymmetry is necessary or relevant in VSA. Thus there is no 'adjunct-argument' distinction in syntax at all, but merely interpretive differences that arise on account of Firstness. Adjunctions can be described as Merge operations that do not result in domain setting or domain assignment. Compare Hornstein & Nunes (2008), where adjunction is formally distinguished in the syntax from other structure-building, based on a difference in how structures come to be labeled. Since Hornstein & Nunes treat Merge as two operations, one of which (concatenation) can apply without the other (labeling), it is not clear how their approach could be squared with KF reasoning.

  Notice also that I have included c-command of the head H as well as H_D, which is to distinguish cases where an operator Merges to H_D by 'tucking in' such that H c-commands the operator. It is not obvious that this possibility should be excluded on empirical grounds, but I exclude it here to simplify presentation of the theory.

to serve as a particular statement of Derivational Drag.[15]

(14) *Firstness*

Interpretive relations in a derivation are uniquely assigned at the first point in a derivation where they are viable.

The mechanisms for domain setting and assignment meet the desideratum of being structurally identical for argument prominence relations and non-argument prominence relations. For example, the definition of 'external argument' can now be thought of as the assignment of v^*_D to X, which arises when X is the first c-commander of v^* and v^*_D (where v^* is 'little v'). Thus when *John* in the sentence *John hit Bill* merges to [v^* [*hit Bill*]], *John* is assigned v^*_D, and as a result *John* is an argument more prominent than anything v^*_D. The relation between v^* and the verb it selects will not be explored here, but the theory of domain assignment is consistent with the view that V raises to v^*, adjoining to v^* at the point in the derivation where v^* and v^*_D are immediately dominated by the undominated node. Subsequent merger of the external argument to the undominated node to [[V v^*] v^*_D] will still be the first constituent to c-command both v^* and v^*_D, and thereby will meet the definition of domain assignment (for assumptions about head-to-head adjunction, see sections 5.1 and 7.2).

The nature of H_D is determined by the properties (e.g., features) of H. Setting of a domain may or may not involve sensitivity to a label in that domain. As mentioned earlier, the theory of c-selection requires that the selecting head find a particular label within its domain if selection is to be successful. However, heads that normally agree with a label in their domain are sometimes permitted a default form if their domain is empty of such a label, a situation without parallel in complement selection. Heads setting scopal domains do not appear to have any relation with a node in their domain, but the burden of successful interpretation is then on whether or not a scopal element, which requires a domain, has the right domain assigned to it. In addition, it is argued in section 4 that some domain-setting heads also assign prominence ordering to multiple labels in their domain where there is no intervention between them. Domain setting in (13a) is thus the most general statement of domain setting, in that it

---

[15]   Firstness, along asymmetric c-command in (5) and the assumptions about domain assignment in (13), will derive much of what Collins (2002) introduces his Locus Principle, insofar as assignments occur as soon as they can, but without the problem of relativized intervention pointed out by Seely (2006). Interventions stated on the heads that introduce domains (enabling assignments) are relativized across different sorts of relations, such that c-selecting heads do not intervene between a probe and goal for agreement relations.

Starke (2004) (see also Jayaseelan 2008) has suggested that specifiers are unnecessary insofar as they are so often in complementary distribution with the heads they correspond to, hence only one of the two is really necessary to satisfy a fixed sequence or template of functional projections. However, heads and the elements that they set domains for in VSA are not always in complementary distribution. Moreover, insofar as Starke assumes that the functional sequence is fixed as a template, rather than as recursive selection as permitted by VSA, a new linguistically specific device, the functional sequence, is added to the theory along with Merge and whatever else is needed to generate the full range of structures. This appears to be a departure from KF reasoning and in spite of some interesting issues that are raised, I will not explore it further.

does not require selection for a label and, moreover, it does not specify what sorts of interveners limit the domain, apart from the hope that (at least the weak form of) the natural intervention hypothesis (10) will provide a leading strategy. It remains to be seen whether the properties assigned to heads in the varieties of domain-setting relations provide sufficient descriptive power for both lexical and crosslinguistic variation, but this power appears to be ample.

3.3.   VSA and the KF

Given this design, most instances of Merge in a derivation, whether they are internal or external, add unalterable information to interpretation, either by introducing a head that sets a domain, or one that orders the labels in its domain for prominence, or else by merging a term for which a domain is assigned. Interpretation proceeding in tandem with tree construction thus initiates considerable derivational drag. The reach of a head setting a domain is limited by the intervention of a more locally c-commanding head, though I have touched on this only lightly so far. That is the essence of VSA design in a nutshell.

How well do these assumptions so far respect the KF? Insofar as domain setting and domain assignment respect c-command, I have not compromised the Mapping Principle for these notions, which are central to all that follows. I am, however, committed to some pre-linguistic relation between certain kinds of potentially discrete notional content and terms that cover those contents. For example, if there is a pre-linguistic notion of an eating event and as a condition of it that something must be consumed, that is, something must undergo that action, then there is perhaps a pre-linguistic precursor to Agent-Patient relations. Such relations can become both more intricate and more generalizable when the Mapping Principle provides for viable expression of them, for example, in c-command structures that map argument prominence in a consistent way.[16]

4.   Case Prominence

One of the practical functions of Agree that has been appealed to within

---

[16]   Although I assume that features enter structure by virtue of the contributions of non-syntactic components, I do not explore here how lexical items bearing labels and features (phonological, semantic, morphological) come to have the labels and features they bear. One of the most dramatic differences between primates that have been taught sign vocabularies and human lexicons is that human lexicons are enormous by comparison and can be casually and productively extended. This difference should, in principle, also result from the KF, at least as a research strategy. There is reason to believe that structure internal of words is interpretively complex. The question is then whether or not it is the advent of the Mapping Principle that makes the lexicon fully viable. Whether one has the view that morphological structure *is* syntax, or that there is an independent, pre-keystone lexical component transformed by access to structure provided by the KF (e.g., being able to map meaningful relations onto word internal c-command found in rich word-internal representations), a theory consistent with VSA is possible. For a variety of recent theoretical approaches exploring the textured internal syntax of words, see Borer (2005a, 2005b), Ackema & Neeleman (2004), Ramchand (2009), amongst others. The exploration of this interesting question is beyond the scope of this essay.

minimalist architectures is to identify constituents that can be moved to A-positions (as in Chomsky 2007a: 25). In a theory without operational triggers that distinguish iMerge from eMerge, an alternative mechanism must indirectly determine which constituents participate in successful A-movement and which constituents do not. In this section I exploit recent developments in the theory of Case assignment and agreement stemming from work by Marantz (1991), Bittner & Hale (1996), and, especially, Bobaljik (2008), who suggest Case assignment arises from a mapping of Case arrays onto prominence relations (as determined by c-command). This independently motivated theory will not only permit us to dispense with triggers for A-movement, but it also turns out to derive syntactic distinction between A-movement and A'-movement, a consequence I postpone until section 5.

In a prominence-mapping (P-M) theory of Case assignment, Case-assigning heads can assign a single Case or a Case array—practically speaking, two Cases, where one of the Cases is more prominent than the other one. The most prominent Case is mapped onto the most prominent argument in the domain of the Case assigner ($CASE_D$), and the less prominent Case is assigned to the next most prominent argument in $CASE_D$. If there is only one nominal, K, in $CASE_D$, then the most prominent Case in the array of the Case assigner is assigned to K. If there is no nominal in $CASE_D$ to assign Case to, none is assigned.[17] A Case sometimes described as 'unmarked' is assigned to the most prominent nominal in $CASE_D$ and a marked Case is assigned to the nominal(s) of lesser prominence in $CASE_D$. No marked Case can be assigned Case again or a conflict I will call 'Case Clash' arises. This means that a marked Case cannot move into the $CASE_D$ of another Case assigner. A nominal with an unmarked Case, however, is still eligible to be assigned another Case. If the unmarked Case is identified as Nominative, for example, then only Nominatives can move into a higher Case domain and be reassigned Case. This is a form of derivational drag, where marked Case assignment has the effect of freezing nominals in place (not literally—movement is free, but Case Clash is a failure at the morphological interface). Only the most prominent nominals in $CASE_D$, when these are Nominative, can be promoted into the $CASE_D$ of a higher Case assigner.

Part of the independent appeal of a P-M Case theory is that it derives Burzio's Generalization, which is the observation that verbs that do not assign an external argument do not assign (structural) Accusative Case. In a P-M theory, this is simply due to the existence of only one eligible argument on the prominence scale.

P-M Case assignment also provides the basis for an elegant theory of finite verb agreement based on the idea that the target of verb agreement is the nearest nominal with an unmarked Case. In English, T assigns the array N-A, with Nominative (the unmarked Case) mapped onto the highest nominal. Empirically, agreement is thus typically with the subject of a transitive verb, but when a verb is intransitive, only Nominative is assigned and agreement with the intransitive

---

[17] This is one respect in which the Case-assigning ability of T is not like an uninterpretable feature, in that it does not cause ill-formedness if it is not assigned (a parallel with Bošković 2007). The Case-assigning potential is simply inert when there is no mapping onto nominals in its domain. The same is true of agreement.

verb argument is the result, regardless of whether or not the single argument is a complement or a subject, a matter elaborated on in section 4.3 for English *there* constructions.

If we add the assumption (as do other P-M Case theorists) that Oblique Cases are invisible to prominence marking, and thus invisible to Case Clash, then it will also be possible for verbs with two arguments, one Oblique, to assign just one Case, Nominative. In languages like Icelandic, there is a richer variety than in English of oblique Cases that are lexical or inherent. As a result, verbs that assign lexical Case to their most prominent argument, will leave only one argument for Nominative Case to be mapped onto. If we assume that T assigns the Nominative-Accusative array and that C assigns Nominative, then a Dative argument in Icelandic, immune to Case assignment by C (and hence immune to Case Clash), can then satisfy EPP (see next section). In (15), for example (from Jónsson 1996: 143), the experiencer argument is lexically determined to be Dative, so the verb complement *Þessir sokkar* gets Nominative Case. T agrees with the highest Nominative, which in this instance appears to be the less prominent argument.

(15) Jóni   likuðu   Þessir   sokkar.   *Icelandic*
      Jon.DAT   *like.PL*   *these*   *socks.NOM*
      'Jon likes these socks.'

There are a variety of more complex Case-marking situations in Icelandic, but this theory follows the logic of Bobaljik (2008), and can claim the same sorts of empirical successes and problematic cases given the basic concordance with principles presented here (see, for example, Bobaljik's account of defective intervention and partial agreement).

A particularly attractive result of the P-M Case theory is that it permits, in Bobaljik's rendering of it, a unification of the rule for subject-verb agreement across languages that have Nominative-Accusative (N-A) Case patterns and languages that have Ergative-Absolutive (E-A) Case patterns. If the unmarked Case in E-A languages is Absolutive, then agreement is predicted to be with the less prominent of two arguments for typical active transitive verbs. Bobaljik is not specific about how the E-A array is assigned, but, in the spirit of Marantz' analysis, suppose that the mapping onto argument prominence is inverted, such that Absolutive is mapped first onto the argument assigned the lowest prominence, and then Ergative is mapped onto the more prominent argument, if there is one. For the sake of discussion, let us assume that either T or Aspect assigns the E-A array and that the Ergative Case has the status of an Oblique Case for subsequent assignment (it is immune to Case Clash).[18]

The charm of this account is especially evident for split Ergative languages

---

[18]   I am aware that this brief determination that Ergative is a structural Case does not do justice to the literature, see for example, Legate (2008: 58) where a number of sources are cited in favor of the view that Ergative is an inherent Case. For the purposes of VSA, treating Ergative as inherent would simply make the E-A pattern similar to the Icelandic Dat-Nom pattern, but more general. Although my position on these matters is not crucial to VSA, it is notable that in split Ergative languages like Hindi, the same argument that is Nominative in one aspect but becomes Ergative in another, which suggests that the term 'inherent Case' is atypically applied to Ergative in such languages.

like Hindi, where only imperfective clauses show the E-A pattern. Suppose that a perfective aspect node can intervene (or merge with T) and count as an inverse Case mapping head, resulting in Erg >> Abs mapping to prominence, as in (16a) (examples from Bobaljik 2008).

(16) a.   Raam–ne   RoTii   khaayii   thii.   *Hindi*
            Ram–ERG   bread–Ø.FEM   eat.PERF.FEM   be.PST.FEM
            'Ram had eaten bread.'

      b.   Siitaa   kelaa   khaatii   thii.
            Sita–Ø.FEM   banana–Ø.MASC   eat.IMP.FEM   be.PST.FEM
            'Sita (habitually) ate bananas.'

      c.   Siita–ko   larke   pasand   the.
            Sita–DAT   boys–Ø   like   be.PST.MASC.PL
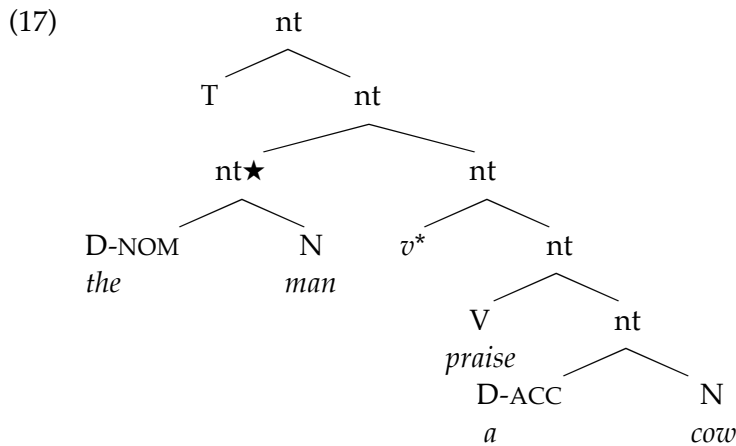   'Sita likes the boys.'

Where the verb is not perfective, the Nom >> Acc pattern remains the default as in (16b). If the subject is Oblique for any other reason (i.e. where the subject is Oblique but not Ergative), then the subject is invisible for T mapping Nom >> Acc, and the object gets unmarked Case (Nominative), as in (16c), just as in Icelandic. Agreement follows the unmarked Case, as indicated by the plural marking on the verb in (16c).

These results suggest that there is strong independent motivation for a P-M Case theory. The rest of this section develops a particular instantiation of a P-M Case theory that serves the goals of VSA.

4.1.   Case Prominence in VSA

Adapting a P-M Case theory to VSA requires some non-trivial adjustments, but none that violate the core ideas of such theories.

For example, since there is no projection of categorial features in VSA, hence no DPs, it is necessary to be more precise about the description, 'nominals in the domain of a Case-assigning head'. The unit assigned Case is D, and so a Case-assigner must rank all of the Ds in its domain for prominence. Intervention will block access to a D that is in the domain of any other Case-assigning intervener, such as C, T, P, or another D. In (17), which diagrams a point in the derivation before *the man* merges above T (about which, more later), the Case domain of T includes two Ds (*the* and *a*), such that neither of the two Ds c-commands the other. Non-terminal nodes are simply marked 'nt'.

(17)

```
                              nt
                      ╱            ╲
                   T                 nt
                              ╱            ╲
                          nt★                 nt
                      ╱        ╲           ╱        ╲
                 D-NOM         N       v*            nt
                 the          man              ╱        ╲
                                            V              nt
                                          praise        ╱      ╲
                                                    D-ACC        N
                                                      a          cow
```

Yet T must rank *the* higher than *a*. The ranking proceeds as follows: T ranks D_A >> D_B if, within the domain of T, there is a node above D_A (illustratively marked with ★), that c-commands D_B. This will usually insure that what is commonly called DP will be the unit that contains D_A and c-commands D_B, and the ranking will proceed accordingly within the Case$_D$ of T. The non-Case-assigning heads between T and D_2 (*a*), such as v^* and V do not intervene for the Case$_D$ of T. The highest ranked D is then assigned NOM(inative) and the phi-features of that D-NOM determine the shape of agreement on T. In more richly inflected languages where N bears a Case that matches its determiner, I assume that N gets that assignment by virtue of being in the domain of D at PF, where the morphological Case assignments are spelled out. In this sense D is also a Case-assigner, and as such is an intervener blocking access to its domain by any higher Case-assigning head.

The core innovation of this Case system now rests on the assumption that any D assigned Nominative by a Case-assigning head can move to a higher domain and still be eligible for a replacement Case assignment, but a D assigned some other structural Case will accumulate Case assignments and be ruled out by the morphological filter in (18).[19]

(18)   *Case Clash*

At the point of linearization, a D must not bear more than one Case.

I make four further assumptions about Case and one concerning EPP.

Oblique Case, once assigned (and by whatever assigns it), is invisible to Case prominence. Thus when T has two Ds in its Case$_D$, A and B, and A is more prominent in the argument structure of the verb, but has already been marked

---

[19]   Languages with more than one Case affix on D are unexpected if (18) is universal, or at least, inflexible, but (18) is essentially a morphological filter, and may plausibly be flexible with respect to language-specific morphological properties. Other apparent counter-examples to (18) as a universal include Case-attraction phenomena, whereby a *wh*-phrase in a relative clause structure bears a different Case from the one it would receive *in situ*, even if that Case is ACC(usative). These effects raise issues for most Case theories, and so I will not explore the possibilities here. Notice that I do not assume a general condition that D must bear a Case, an issue that arises in section 7.

Dative, then only B is visible for Case prominence. This is just the VSA instantiation of the idea already introduced in P-M Case theory to account for agreement patterns like those in Icelandic.

All proper names that receive Case occur with (sometimes null) determiners. This is a theory-internal requirement, since I assume that N cannot be ranked for Case prominence, but determiners do co-occur with proper names in the world's languages, for example, in German dialects and Greek, and the existence D in nominals containing proper names has also been argued for by Longobardi (1994).

The generalization about 'unmarked' Case is probably more neutrally stated as 'reassignable' Case, or R-Case, since Nominative clearly has a marker in many languages (e.g., Icelandic). While this is often true of Nominative and Absolutive marking in the world's languages, it is by no means fully general, and so I will leave the morphological Spell-Out of R-Case to language-specific morphology (see Legate 2008, for a similar conclusion, although in a non-PM Case theory). Thus I would restate Bobaljik's agreement proposal as follows:

(19)   Finite T agrees with highest R-Case in its domain.[20]

The nullity of PRO is not derived from Case assignment under prominence nor by the absence of Case; rather PRO is Case-marked and, even in control structures, bears a Case according to its context, just as other nominals do (see Landau 2006: 154-157 and references cited there). I assume that PRO is assigned Case by the C that introduces infinitives, unless the EPP is satisfied by an Oblique (as some of the concord phenomena with the Case of PRO show in the reference cited).[21]

Finally, I assume the EPP as in (20):

(20)   *EPP*

T$_D$ must be assigned to a non-terminal.

This stipulation is no more conceptually attractive than several others in the

---

[20]   I abstract away from those languages where a verb agrees with more than one argument, as such cases are orthogonal to the discussion of VSA instantiations.

[21]   Since the demise of the PRO theorem with the theory of government, no interesting alternative to account for the nullity of PRO has emerged, other than Hornstein's (1999) movement approach, which remains controversial (see, amongst others, Landau 2000, 2003, Hornstein 2003, Boeckx & Hornstein 2006, and van Urk, forthcoming) and which I will not address here. I will not attempt to provide a theory why PRO is null in this essay. For further critique of the place of PRO in traditional Case theory, see McFadden (2004) and references cited there.
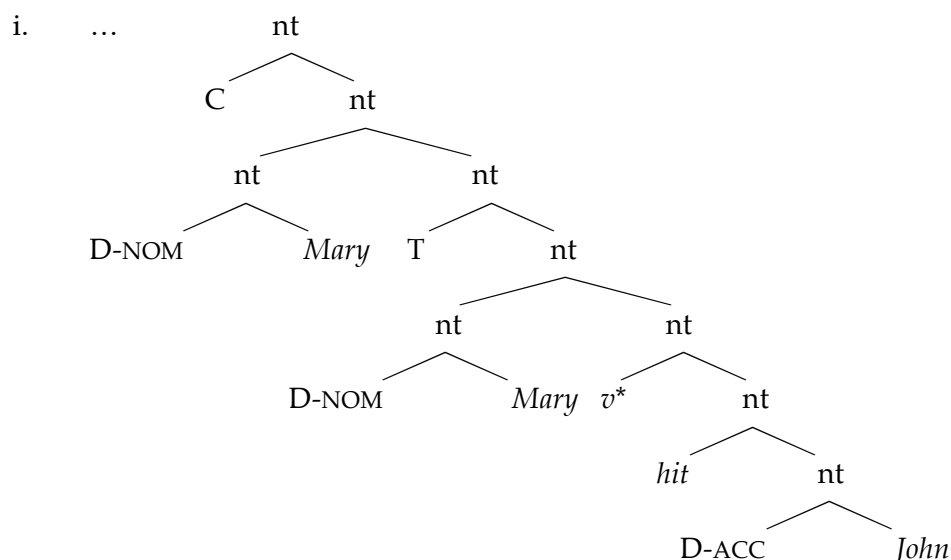
In so far as the Case assigned to PRO is not realized in morphology, for whatever reason, I depart further from Bobaljik's (2008) view that Case assignment is assignment of 'm-Case', a form of Case that always has a phonological form. I have to assume that what morphological Spell-Out sees is somewhat more abstract, as I do for R-Case, but Case is not the key regulator of movement for passive or raising (or the distribution of PRO) as it is in GB-era abstract Case theories. Although I have not compared this account to that of Legate (2008) in any detail, my position on the relation to M-case is more in line with her position than with Bobaljik's.

literature, insofar as it involves an interpretive assignment with no interpretive content, but it does eliminate the last justification for an operational trigger for movement. As I will demonstrate below, this way of formulating the EPP will avoid many of the complications that other theories are forced to without introducing counter-cyclic instances of Merge (as are found, for example, in Chomsky 2007a, 2008, where T must inherit features from C before it can attract a constituent to its Spec-position) and without introducing features that are crucially uninterpretable in some other component. The treatment of EPP in (20) holds out the hope that the right theory of 'subjects' may yet reduce EPP to a contentful form of domain assignment, such as scope or argument prominence, perhaps as Rizzi (2006) has suggested (see the discussion of criterial positions in section 5).

As a means of fleshing these ideas out as they apply in specific cases, consider the annotated derivation in (21a–h) for *Mary hit John* and the resulting structure in (21i). In contrast to Chomsky (1995, 2001), I assume that there is no numeration specifying the input to structure-building because the only arbiter of what constitutes a well-formed output is whether or not the result of the derivation is semantically and phonologically interpretable at the interfaces.

(21)   Mary hit John.

   a.   By hypothesis, D must merge with the name *John*. Similarly, a D merges with *Mary*.

   b.   If *hit* is merged with [D *John*], then [D *John*] is the selected domain for *hit*, hit_D, irrevocably formed at this point in the derivation because *hit* selects for D.

   c.   If v^* is Merged to [*hit* [D *John*]], then v^*_D is [*hit* [D *John*]] because v^* selects for V.

   d.   If [D *Mary*] is Merged to [v^* [*hit* [D *John*]]], then [D *Mary*] first c-commands v^*_D (and v^*) and [D *Mary*] is assigned v^*_D (i.e. [D *Mary*] is the external argument of *hit*).

   e.   If T is Merged to [[D *Mary*] [v^* [*hit* [D *John*]]]] and T selects for $v(^*)$, then [[D *Mary*] [v^* [*hit* [D *John*]]]] is T_D. Since T sets prominence for Case in English NOM >> ACC, and the D of [D *Mary*] is more prominent than the D of [D *John*] in the $CASE_D$ of T, [D *John*] will be ACC and [D *Mary*] will be NOM.

   f.   EPP requires that a non-terminal must be assigned T_D and since [D *Mary*] has R-Case (Nominative), it is the only candidate to first c-command T and T_D by iMerge.

   g.   When C Merges to [[D *Mary*] [T [[D *Mary*] [v^* [*hit* [D *John*]]]]]], the $CASE_D$ of C includes the higher [D *Mary*] only, since T intervenes for the rest, and the [D Mary] is re-assigned NOM by C. If [D John] had satisfied EPP, it would fail by Case Clash.

   h.   The output: [C [[D *Mary*] [T [[D *Mary*] [v^* [*hit* [D *John*]]]]]]]

i.   …

```
                          nt
                    ┌──────┴──────┐
                    C             nt
                            ┌─────┴─────┐
                           nt           nt
                       ┌────┴────┐   ┌───┴────┐
                    D-NOM     Mary   T        nt
                                          ┌────┴────┐
                                         nt         nt
                                     ┌────┴────┐  ┌──┴───┐
                                  D-NOM    Mary  v*      nt
                                                     ┌───┴───┐
                                                    hit      nt
                                                          ┌───┴───┐
                                                        D-ACC    John
```

I assume for now that, given multiple occurrences, only the highest is pronounced (see section 6), thus *Mary* is pronounced outside of T. A few comments are necessary here concerning C. I am assuming first that T is always selected (i.e. by C or a lexical head), as seems necessary to say independently of any considerations peculiar to this theory. Notice that either iMerge or eMerge could satisfy EPP, but for reasons that remain murky for English, transitive expletive constructions are not allowed, and so *there* cannot be inserted at step (21f) (a problem shared by most other theories). In the absence of an eMerge option in this instance, movement to subject is forced by the domain assignment condition embodied in EPP. As remarked above, [D *Mary*] is marked NOM twice, but without Case Clash, because one Nom assignment (by C) simply replaces the other (by T). If [D *John*] had iMerged to first c-command T_D at step (21f), then [D *John*] would induce Case Clash, since ACC on [D *John*] cannot be reassigned any other Case (including another ACC).

    The same system derives the sentence in (22), an instance of raising. Notice that I assume that infinitival T, like tensed T, assigns Nominative Case and, as in standard accounts, that there is no C in English raising infinitives.

(22)   Several men appear to be leaning on the balcony.

    a.   The first steps include Merge of *balcony* and *the*, where *the* selects N, then Merge of *on* with *the balcony*, where *on* selects D. Since *on* is a P, it is a Case assigner and assigns Oblique to the most prominent (only) D in its domain which is *the*.

    b.   Merge of *leaning* (I ignore *–ing* here) to *on the balcony*, where *lean* selects P and takes *on the balcony* as its domain, depending on whether or not *on* is an intervener, since *the balcony* is also the domain of *on*. Merge of *v** then takes *leaning on the balcony* as its domain and Merge of (previously formed) *several men* fills the external argument (EA) slot for *v** (i.e. *several men* is assigned *v*_D*).

    c.   *Be* is a V that Merges to [[*several men*] [*v* leaning on the balcony*]],

setting the contents of the latter constituent as *be*$_D$.

d. T (*to*) Merges to [*be several men leaning on the balcony*] selecting the latter as its domain up to intervention. T detects the prominence relations in its domain, but since *on* is a Case assigner that intervenes between T and *the*, T orders only *several* and assigns it Nominative. Infinitival T lacks agreement features.

e. [*Several men*] iMerges to [*to be several men leaning on the balcony*]. This step is derivationally optional but necessary to satisfy the interpretive condition EPP.

f. Then *appear* Merges to [*several men* [*to* [*be several men leaning on the balcony*]]], and then *v* Merges to [*appear several men to be several men leaning on the balcony*].

g. When [+tense] T Merges to [*v appear several men to be several men leaning on the balcony*], two things happen. Since T is a Case assigner, it ranks all the Ds in its CASE$_D$. Since subordinate *to* is also a Case assigner, it intervenes, and so the only D in CASE$_D$ of [+tense] T is the higher occurrence of *several men* which is not harmed by being reassigned Nominative. Also, [+tense T] has agreement features, so it agrees with the highest Nominative in its agreement domain, *several men*.

h. Then *several men* iMerges to [T v *appear several men to be several men in the room*]. This satisfies EPP.

i. C merges to [*several men appear several men to be several men in the room*] because [+tense] T is always selected by C. Since step (h) moved *several men* into its domain, C orders *several* as its most prominent D in CASE$_D$ of C and reassigns it Nominative.

'Pronounce highest occurrence' (see (52) in section 6 below) will derive the right phonology, although I will have more to say about which of multiple occurrences are pro-nounced in later sections.

One major advantage of this account over many others in the literature is that it is never necessary to assume any counter-cyclic movement. Chomsky (2008), for example, requires that movement to matrix [Spec,TP] position is only triggered *after* C is merged to TP, the features of C are inherited by T, and then iMerge is triggered from the lowest position of *several men* to the highest position of *several men*, but stopping in the lowest [Spec,TP] to satisfy EPP and then in the highest [Spec,TP]. No stutter-step anti-cyclic movement is required in the derivation of (22).

In the next two subsections, further mechanisms for, and consequences of, mapping Case and agreement onto prominence relations will be fleshed out.

4.2. *Case Prominence and Derivational Drag*

I assume that C, T, P, and D are potential Case assigning heads and that the intervention of any one of these heads blocks Case prominence ordering and assignment by any higher head. Lexical or inherent Oblique cases are not assigned

prominence or Case by the Case-assigning heads, such that a nominal K without lexical or inherent Case will count as the most Case-prominent nominal in its domain even if a nominal J with inherent or lexical Case c-commands K in the domain of K's Case assigner. Any nominal that is assigned more than one Case without reassignment will be excluded by Case Clash (as stated in (18)). In languages where Case is more frequently inherent or lexical, hence immune to Case Clash under reassignment (because they won't be prominence-ordered for Case assignment by a head), nominals with non-structural Case may move to higher Case domains with impunity. Case assignment by a Case-assigning head is obligatory unless there are no eligible D nodes in its domain.[22]

As mentioned with respect to PRO, I assume that infinitival C also assigns Case. One way of blocking raising from the position of PRO is simply to assume that infinitival C always assigns Accusative, whether it is null C or English *for*. On this account, (23b) and (24b) are excluded for the same reason, namely, iMerge of *him* in a domain of higher Case assignment will result in Case Clash.

(23) a. It is important for him to leave.
      b. * *He* is important [for *him* to leave].

(24) a. It is important [C_{NULL} PRO to leave].
      b. * *He* is important [C_{NULL} *him* to leave].

It is instructive to see how this reasoning applies to passive structures. In English, when a transitive verb like *praise* is in its passive form, it takes no external argument (suppressed by a rule applying to the lexical entry) and is selected by v, which, unlike v^*, does not select a domain that can be assigned to an external argument. For the moment, let us assume that no argument is assigned to v_D, a result to be derived momentarily. As a result, there is only one nominal in the domain of T, the direct object, and that nominal gets R-Case. The nominal with R-Case is the only one eligible to be reassigned Case by C, and so it is that argument that satisfies EPP by being assigned T_D.[23] Where a locative expletive (*there*) is available to satisfy EPP, the object may remain *in situ* and receive NOM as the only visible nominal visible to Case assignment in CASE$_D$ of T.[24]

At this point we have seen that there is a derivational drag effect on arguments assigned structural Case that is not R-Case, such that only an R-Case nominal can move to a higher Case domain.[25] The effects of this restriction have been shown to be positive, insuring that all non-R-Case structural cases are frozen in place, while permitting only R-Case nominals, or Oblique Case nominals, to move to higher Case domains. However, even A'-movement of non-R-Case nominals that have structural Case is also apparently ruled out in this system, a matter to which I return in section 5.

Another interesting derivational drag effect is induced by these assumptions about Case assignment. If a nominal moves within in $CASE_D$, it will move to a position that c-commands the copy of it that is in its point of origin. In such instances, both nominal positions will be evaluated for prominence in Case, with the result that two occurrences (OCCs) will have different Case assignments, where the lower one cannot be reassigned. This will result in Case Clash.

(25) $[H_C [OCC … [… OCC …]]]$

---

[25]   Prenominal Genitive nominals (PGNs) are generally assumed to c-command the D of the nominal they are contained in (e.g., they are assumed to be in '[Spec,DP]'), which means the D of the PGN is in the $Case_D$ of the Case assigning head external to the nominal. That Case-assigning head should determine the prominence prenominal D >> containing D, resulting in the wrong Case array (e.g., for a D that should be Nominative, such as *mother* in *John's mother laughed*) and the Genitive D could also be interpreted as a potential complement by a selecting head.

The Case problem disappears if Genitive is an oblique Case, but Genitive appears also to be structurally assigned. Since Spec-Head agreement is not available in VSA, the only way to model Genitive assignment in this theory is to posit a Gen-assigning head H-Gen that c-commands and Case-assigns the D of the PGN. If H-Gen is itself a determiner (one that can bear a different Case than it assigns to a D in its domain) then PGNs must originate below H-Gen. The following contrast suggests that some PGNs are below H-Gen and others are above it.

   (i)   What kind of bread was she willing to buy a loaf of?

   (ii)   * What kind of bread was she willing to buy John's loaf of?

   (iii)   Which of his latest escapades was she willing to hear an explanation of?

   (iv)   ? Which of his latest escapades was she willing to hear John's explanation of?

It is pointed out in Safir (1987) that extraction out of nominals that have PGNs is improved if the PGN bears a thematic relation to the head N. In the latter situation, it appears that the PGN is optionally assigned a theta-role by merger above $[N N_D]$ (or perhaps a nominal version of v^*). Thematic roles undetermined by the head N seem to be assigned a default theta-role by merger above $[H\text{-}Gen\ H\text{-}Gen_D]$. If so, the PGN initially emerges below H-Gen where it receive Case, but the PGNs not thematically related to the head N must raise to receive a theta role above H-Gen. High PGNs block extraction from the full nominals more completely, which would be explicable if the left periphery of the nominal domain is an escape hatch for extraction. This account requires a theory of low PGNs, an analysis, which happens to be consistent with VSA tenants of Case assignment. Moreover, the mechanisms described insure that PGNs are already theta-assigned before a higher selector can assign any theta-role to them, and perhaps in certain contexts more than one thematic assignment is possible (e.g., possessor raising constructions). A more fleshed out version of this proposal is too large for discussion here.

On this account, it is impossible for any nominal in the Case domain of a Case-assigning head, H_C, to move to another position within the domain of H_C. This will prevent iMerge of the direct object just above [v v_D] in passive structures, or else the configuration in (25) will arise within T_D. This would appear to rule out a great deal of possible movement, perhaps too much, especially in scrambling languages. I return to this issue in section 5, where the relation of A'-movement to Case assignment is examined.

4.3.   Relativized Domains: Case and Agreement

Chomsky (2008: 154) contends that

> [uninterpretable features] must therefore be valued at the stage in com-putation where they are transferred—by definition, at the phase level—and this must be the same stage for both transfer operations, again supporting the optimal assumption that transfer to both interfaces is at the same stage of derivation.

This assumption is not necessary, it is not obvious that it is the optimal assumption, and, moreover, some of the stipulations required to support this hypothesis suggest it is misconceived. For example, Chomsky appeals to EPP features in raising structures to attract movement within a phase to potentially many subject positions, yet Agree only applies between the highest T and an argument embedded potentially several TPs below. This difference between attraction by EPP features and Agree suggests that phases are just cycles of a particular kind, not the *only* cycles. Rather it would appear that cycles of different kinds, and, in VSA, interveners of different kinds, are *relativized* to the sort of relation that they establish. In other words, the notion that there are windows (domains) in the course of a derivation when certain relations can or must be established is preserved, but it is not assumed to be the same window for all relations.[26]

Relativized intervention can be illustrated with the difference between Case assignment and agreement as modeled in VSA. For example, a Case-assigning node H will set a domain $CASE_D$, its sister, which includes all the nominals in $CASE_D$ outside the next lower Case-assigning head. If there are two nominals, then they will be ranked for prominence (ordered) and aligned with a Case array that is determined by H. $CASE_D$ for T_1 is everything dominated by its sister up to intervention, i.e. the $CASE_D$ for T_1 does not extend any lower than the sister to the next Case-assigning head that T_1 c-commands. If the closest Case-assigner to T_1 is another T, T_2, then anything $CASE_D$ for T_2 is not in the ordering that includes those nominals in $CASE_D$ for T_1.

---

26   For arguments that A-movement and Agreement have different domains, see Bobaljik & Wurmbrand (2005), and for an argument that the domain of Principle A is not congruent to the domain of A-movement, see Safir (2004a: 147-156). Seely (2006: 202) points out that the domains for selection and Agree are different. Bošković (2007) argues in particular that phases and the domain of agreement are different, and that A'-movement does not involve intermediate triggers for a constituent which has features not satisfied by any structure in the tree, but could involve Agree applying across a larger domain when the probe is introduced.

We have seen how this works for simple cases like (21) where T assigns the Case array NOM >> ACC and that T agrees with the highest NOM in its domain, namely the subject when it is in EA position. In (26), however, agreement is with the nominal within the complement of *be*.

(26)   There are three men missing.

In this example, T orders the only argument in its domain (*three men*) for Case prominence and then assigns it NOM Case. However, instead of raising *three men* to the matrix position to eventually satisfy the EPP requirement of matrix T, *there* has been inserted. Such cases as these recall the Icelandic DAT/NOM pattern, where agreement is with the highest NOM even when it is not a subject. Such a parallel suggests that *there* plays this role because it is locative, and thus an Oblique Case.[27] As in Icelandic, it may then be assumed that *there* satisfies the EPP requirement of matrix T, it is in the Case prominence domain for C, but it is not assigned Case by C. (After all, the use of a locative in subject position with postverbal agreement is also found in English examples like *In this village are found many fine woven goods*.) In other words, I am treating *there* as something like *D.Loc-here*, where locative meaning has been bleached away. This then provides a natural account for the distinction between *there* and Standard English *it*, which is not oblique and which agrees with the verb. However, examples like (26) still correlate Case and agreement, insofar as the highest NOM, the associate of agreement, is within $CASE_D$ of T.

To see how the setting of domains is relativized, consider, for example, ECM structures in English and let us suppose that T_1 is the matrix tense T.*past* which takes everything in its sister non-terminal node as its domain, and that the next Case ordering domain is that of T.*to* of the subordinate clause.

(27)   a.   He T.*past* expected her T.*to* hate them.
       b.   There T.*past* were expected T.*to* be several men in the room.

For (27a), *her* and *them* will be Case-ordered and assigned by T.*to*. The pronoun *her* receives R-Case from T.*to*, but *them* will be marked ACC and thereby frozen in place. When *her* raises up to be assigned the domain of T.*to*, as it must to satisfy EPP, it will then be embedded as *expect* merges to [T.*to* T.*to_D*] and selects T.*to*. When *he* is merged to v^* above *expect*, it becomes the external argument (EA) of *expect* and will be in position to receive R-Case from matrix T.*past* which will order *he* >> *her* for Case prominence and assign ACC to *her*, overwriting R-Case. *He* will then be the first c-commander of T.*past_D* above T.*past* and satisfy the EPP. When C is Merged, *he*, which has R-Case from T.*past* will be reassigned R-Case (NOM) by C.[28]

---

27   It may be necessary to assume that there is a locative head between T and v that assigns a LOC argument to its domain. See Linares (forthcoming) for a proposal along these lines.

28   If the C of tensed clauses assigns R-Case, then in theory, the R-Case subject could raise into the next Case domain. This does not normally happen because movement to a position immediately above C_D is movement to a criterial position, which a non-*wh*-subject is inappropriate for (see section 5.2) and movement to a higher position will not be able to escape

In (27b), however, *several men* is the highest nominal in the lower domain (that of T.*to*) and receives NOM in situ. *There* satisfies the EPP by being the first c-commander of [T.*to* T.*to*$_D$]. Subsequent iMerge of *there* satisfies the EPP requirement for matrix T.*past*. The issue now concerns agreement of T.*past*, which is empirically plural. Insofar as locative *there* counts as an Oblique Cased nominal, neither matrix *there* nor its subordinate copy qualifies as the nearest R-Case (NOM) to T.*past*; rather T.*past* agrees with *several men*, even though *several men* is in the domain of T.*to*. However, in the relativized system assumed here, T.*to* is not an intervener for agreement because it does not bear agreement, i.e. it is not in an agreement relation with anything, so the domain of T.*past* with respect to agreement includes the contents of the lower clause and its highest NOM, which is *several men*.

In a theory where phases are not relativized to relations in this way, it is necessary to assume that *there* carries a plural agreement feature or that the lower phase is invisible selectively or generally in this sort of example. Rather than complicate the derivation in the latter way, intervention is relativized to the relation involved. A natural account of long distance agreement in *there*-sentences with raising is the result.

5.   Criterial Positions and Reducing the A/A′-Distinction

What have been known since Chomsky (1981) as A′-movements, the quintessential example being *wh*-movement, have a different class of properties from A-movements, which seem to revolve around movement to subject(-like) positions. The A′-movements displace constituents in such a way that the determination of scopal properties or information-structure values are almost always involved (although *tough*-movement structures raise questions), as contended most recently and explicitly, for example, by Rizzi (2006), whereas A-movements appear to revolve around Case and agreement relations, *not necessarily* involving scope or information structure (though they often do have effects of this kind). Chomsky (2007a: 25) still distinguishes them as follows: "A-movement is IM contingent on probe by uninterpretable inflectional features, while A′-movement is M driven by [edge features] of P" (where IM = iMerge).

That Case requirements are involved for A-movement, but not for A′-movement, is a distinction that should be effaced or derived. The Case prominence theory permits us to achieve this result, insofar as the only difference between A-movement and A′-movement resides in the different strategies they exploit to avoid Case Clash, not with respect to features that attract movement, nor with respect to the requirements of interpretation they satisfy (e.g., edge features appear to be linked to the satisfaction of interpretive requirements).

Consider simple examples such as (28a), where it is clear that *whose brother* must reach the position where its scope is assigned, which would be the point in

---

forced Spell-Out in domain C$_D$ (see section 6). There are languages where the subject of a tensed clause can raise into the next higher domain and receive ACC (Korean), and for that to be possible, it must be assumed that there is a C (or some other domain-assigning head in the intermediate zone above C) which sets a domain compatible with a non-*wh*-phrase.

the derivation where it iMerges just above C. This it presumably must do in order to achieve the proper assignment for interpretation. However, it is not obvious that the Case theory proposed so far can produce the right result, especially given the schematic tree in (28b).

(28) a. They wonder whose brother John saw.

   b.

```
                nt
               /  \
              C    nt
            they   /  \
                  T    nt
               wonder  /  \
              [whose brother]  nt
                          /  \
                         C    nt
                       John   /  \
                             T    nt
                      [whose brother]  nt
                                   /  \
                                 John  nt
                                      /  \
                                    v*    nt
                                        /  \
                                       V   [whose brother]
                                      saw
```
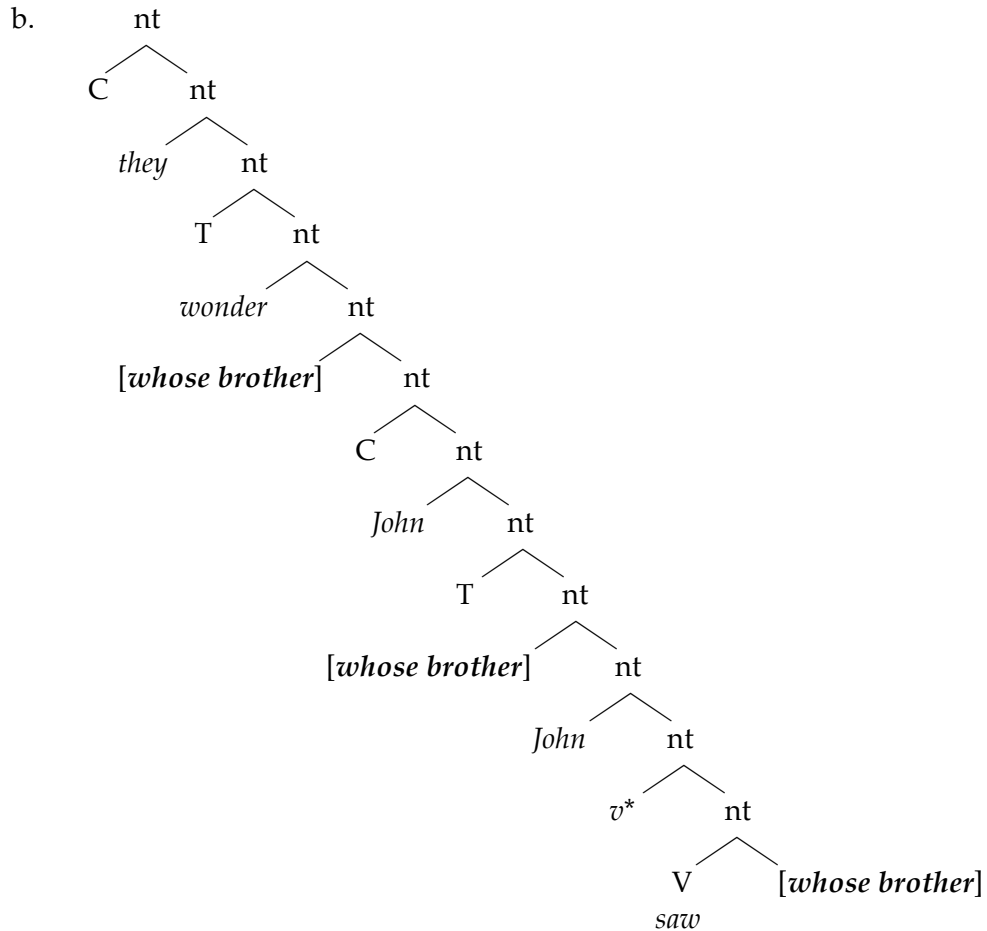
The representation in (28b), where *whose brother* has 'adjoined to *v*P' en route to its final landing site, exposes *whose brother* to Case assignment by T in the lower clause, where the higher occurrence would get NOM and the lower one ACC, leading to Case Clash. Moreover, *John* would receive ACC from subordinate T instead of NOM, and should then induce Case Clash when it moves to subject position to satisfy. Even if we solve the problem below subordinate T, subsequent iMerge of *whose brother* above the embedded C exposes it to Case assignment from matrix T, which would assign it Accusative.

In this section, I will explain how the scope of overt wh-movement is assigned in VSA and how the solution to the problem raised by Case assignment resolves the A/A'-distinction without relying on pair-merge or any other special addition to VSA that is not required by other theories. I postpone to section 6 why it is that A'-movements do not typically take place directly to criterial positions, but must pause below their final destination to satisfy locality conditions associated with cyclicity.

5.1.   *Revising Extension*

Before I present my account of the A/A'-distinction, it is necessary to begin by introducing a revision of Extension which *is* required by other theories, at least in some form. Extension has actually been abandoned in some recent accounts, more typically in practice than by any explicit rejection (e.g., counter-cyclic movement in Chomsky 2008). Although the revision required has been motivated independently of the theory proposed here, it is a revision my subsequent proposals will exploit.

Many accounts of anti-reconstruction effects depend on violations of Extension as stated in (2) (see Lebeaux 1991, Chomsky 1995, Bhatt & Pancheva 2004, Fox 2002, Stepanov 2001, and Safir 2004 for discussion). The variable-binding contrasts below are from Safir (1999: 601–602).

(29) a.   * [Which reviews of every poet's$_i$ book]$_j$ does he$_i$ try to forget t$_j$?

   b.   ?? [Which analysis of every poet's$_i$ book]$_j$ is his$_i$ mother most afraid of t$_j$?
        {answer, for example, 'the Freudian one'}

   c. ? [Which reviews of every poet's$_i$ book]$_j$ t$_j$ give him$_i$ the most satisfaction?

(30) a.   [Which book on every poet's$_i$ shelf]$_j$ is he$_i$ particularly proud of t$_j$?

   b.   [Which book on every poet's$_i$ shelf]$_j$ is his$_i$ mother most proud of t$_j$?
        {'The one dedicated to her'}

   c.   [Which book on every poet's$_i$ shelf]$_j$ t$_j$ gives him$_i$ lasting satisfaction?

The bound reading of *he* in (29a) fails on the assumption that a copy of *which reviews of every poet's book* inhabits the position notated with a trace, hence *he* c-commands its antecedent, leading to a violation. Similarly, a copy in the position of the trace in (30b) would also induce a weak crossover effect, but not if the copy were in subject position (30c). It appears that certain prepositional phrases must be attached before iMerge of the whole *wh*-phrase takes place and these PPs leave a copy behind that results in variable binding violations. The effects disappear in (30), however, as if *on every poet's shelf* were not part of what is left in the trace position (following a line of reasoning from Lebeaux 1991). Safir (1999) proposes that if *on every poet's shelf* is merged to *which book* after *which book* moves, then strong and weak crossover effects are avoided. But merge of *on every poet's shelf* to *which book* after iMerge of *which book* violates Extension because *on every poet's shelf* does not extend the undominated node R, as illustrated schematically in (31b) and in the diagram in (31c).

(31) a.   [$_R$ [*which book*] [is he$_i$ particularly proud of [*which book*]]

   b.   [$_R$ [[*which book*] [on every poet's$_i$ shelf]] [is he$_i$ particularly proud of [*which book*]]

c.
$$nt=R$$

```
                          nt=R
                      /          \
                   nt              nt
                 /    \          /      \
              nt       nt    [is he_i particularly proud of [which book]]
            /   \     /   \
      [which book] [on every poet's_i shelf]
```

[which book]　[on every poet's_i shelf]
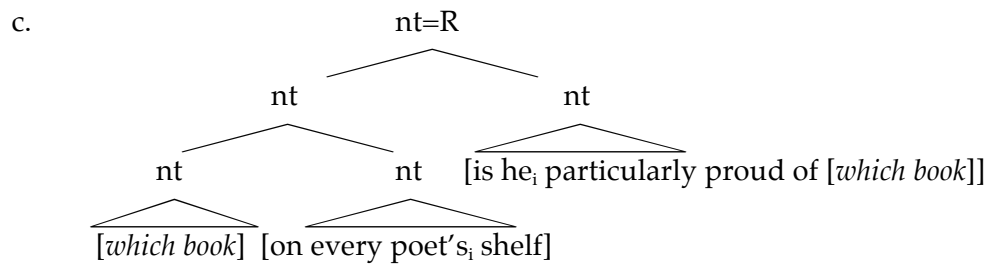
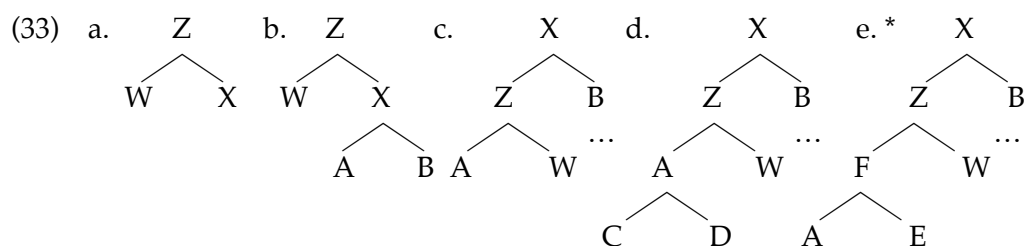Every appeal to late attachment (or 'late merge') in the literature involves an Extension violation of this kind.

It would be unfortunate, however, to jettison Extension (or just stipulate that it does not restrict 'adjunction', as in Chomsky 1995: 189–190) from the theory entirely, as it is the condition that prevents a rich variety of counter-cyclic movements, possibilities that greatly enhance the descriptive power of grammars. It is still possible to capture the idea, however, that Extension is always sensitive to the position of the undominated node, if we provide a more articulated idea of what the top of the tree is.

(32)　*Revised Extension*

After every instance of Merge, M^i, the undominated node of the resulting structure immediately dominates a node it did not immediately dominate before M^i.

The structures in (33a–d) result if W has just been Merged (abstracting away from linear order) and Revised Extension has been respected. This definition exploits the long held assumption that Merge always produces binary branching trees. Cases where both terms of Merge are terminals (themselves undominated nodes), as in (33a) will be immediately dominated by a new node Z, which is the non-terminal formed by M^i. Since Z is new what it immediately dominates is newly dominated by it, and the same account extends to (33b), where one (or both) terms of Merge are non-terminal(s). However, a more novel possibility arises where W adjoins to A, as in (33c) or (33d) (where A is terminal or is not, respectively) in each case creating Z, a node newly immediately dominated by the undominated node X after M^i applies, consistent with (32).[29]

---

[29]　Non-terminal nodes do not bear labels in VSA, so it is fair to ask how we would know that a particular non-terminal node that was not dominated 'before' was *in fact* not dominated before. As a technical matter, suppose that every node created by Merge is assigned a 'term index' and every term of Merge is assigned a new term index only if it does not already have one. In cases where Revised Extension is satisfied by submerging one immediate daughter of the undominated node, but not the other, the undominated node is not a term in that Merge operation; only the nodes that together form a new node are terms in that operation. The term index is then what is copied in iMerge and could then replace the very similarly employed 'numeration index', which will be eliminated in VSA, which has no numerations. The ability to describe copies of both non-terminals as sharing the same term index, just as terminal copies do, would simplify references to copy sets (for terminals and non-terminals), especially in cases where a terminal node has been extracted from a non-terminal copy.

(33) a.　Z　　b.　Z　　c.　　X　　d.　　X　　e. *　X

Tree diagrams:

a. Z dominates W, X

b. Z dominates W, X; X dominates A, B

c. X dominates Z, B (…); Z dominates A, W

d. X dominates Z, B (…); Z dominates A, W; A dominates C, D

e. * X dominates Z, B (…); Z dominates F, W; F dominates A, E

As a matter of useful terminology, let us say that *W subMerges A* in (33b–d) by alienating A from immediate domination by the undominated node, a status A loses after M^i applies. Notice further that Merge could not subsequently apply to some node E to create a new node F immediately dominating A and E in (33b–d) because that would violate Revised Extension: The undominated node X still dominates the same two nodes, Z and B, before and after E is merged.

Revised Extension permits more attachments than Extension in (2), but it still insures that Merge operations grow only the top of the tree.[30] Some instances where nodes are subMerged, however, permit structure-building crucial to a variety of current proposals that are technically excluded in the earlier formulation. For example, both late Merge, described above, and 'tucking in' as proposed by Richards (1999) are instances of (33d). The structure in (33c), where the terminal W subMerges the terminal A, models iMerge of a head to another head, an operation frequently appealed to in the literature, and one long known to be problematic for Extension (e.g., as noted in Chomsky 1995: 327 and Bobaljik & Brown 1997). The latter three possibilities are all excluded by Extension as in (2), but permitted by Revised Extension.[31]

---

[30]　SubMerge does not change any assumptions about derivational c-command (see fn. 6) nor should it be confused with questions about sideward movement. The subMerged node does not fail to c-command any position it c-commanded before subMerge. Sideward movement occurs when one term of Merge, A, is contained in K, and the other term of Merge is J, where J is undominated and does not dominate K. If A Merges to J, then J is expanded by the operation, satisfying Extension in (2), but the resulting iMerged copy of A does not c-command anything in K, including the copy of A in its launching site, as noted by Bobaljik & Brown (1997). Unregulated, sideward movement could be intersentential. Epstein *et al.* (1998: 103) argue that such a movement could not satisfy economy conditions, that is, it could never be triggered. For example, Agree could never hold between a probe and a goal not contained in its complement, but this will not do if iMerge is not triggered, as argued here. In VSA, however, intersentential sideward movement leads to non-viable relations anyway, since the iMerged element L does not c-command its copy L'. Where L and L' are not in a viable relation, nothing would prevent them from receiving independent semantic assignments (e.g., the pronoun *he* could be assigned a value independent of its copy in another constituent). Issues concerning the Spell-Out of multiple copies would still have to be resolved where c-command cannot regulate Spell-Out (see fn. 35).

　　The output of sideward movement could be viable if there is subsequent movement to a position that c-commands both copies, permitting criterial interpretation of the copies in the relevant way. Some accounts of parasitic gaps have this character, but they must rely on chain formation and c-command (e.g. Nunes 2004: 91). For reasons to doubt the theoretical legitimacy of chain formation, see Safir 2008: 345–346).

[31]　It is reasonable to ask how well Revised Extension fits with the KF reasoning. In principle, the revision of Extension costs nothing, because Merge, which Extension restricts, is assumed to be prior to the KF in any case. Just the same, an implicit claim has been made that subMerge as in (33c–d) is part of the pre-linguistic operation that generates syntactic structure, a complication beyond simple Merge in (1), and I have no evidence for this (nor any obvious way to search for such evidence).

5.2.   *Criterial Positions*

Rizzi (2006) calls positions that must be reached for a phrase to be interpreted a *criterial position* for that phrase (along the lines of the *wh*-criterion or the *neg*-criterion, as in the references he cites). He suggests that the range of criterial positions may include focus and topic positions as well as those that participate in scope interactions, along the lines of Beghelli & Stowell (1997) (see section 7). Indeed Rizzi even holds out hope that EPP is a criterial movement, which would be more suitable for the VSA approach and perhaps others, but the evidence for this further extension is currently slim. All of these criterial assignments are instances of (13b) for a particular kind of domain assignment. In this section, I explore (13b) as it relates to assignment of scope for *wh*-Q-phrases.[32]

(34)   *Assigning Scope for* Wh-*Q Phrase*

The scope of *wh*-Q is assigned when *wh*-Q is assigned a domain that matches its quantificational features/properties.

I assume that a questioned constituent, the *wh*-Q-phrase, can only receive a scope consistent with its interpretation if it is assigned WH_D which is set by C_{WH}, so unless this configuration is achieved, or some other way to interpret the *wh*-Q-phrase is introduced, then the result is a failure of interpretation.

(35)   *Domain Assignment Failure*

An element Y that must be assigned a domain of type X is not assigned a domain of type X.

Thus, Domain Assignment Failure occurs when no domain is assigned to an element that needs one or the domain that is assigned does not match the element it is assigned to, that is, it is a form of incoherence. IMerge is optional in VSA, not syntactically triggered, so cyclic movement of a *wh*-Q-phrase is optional even if it is to a non-criterial position, as it is when it is Merged above *v* (higher than the external argument) or above C_{THAT}, which does not set a scopal domain. Thus no intermediate triggers are needed for long cyclic A'-movement, but failure of the phrase to move cyclically does not lead to an interpretable outcome.[33]

There is an important consequence of this formulation of domain setting and assignment for cases of head to head movement frequently found in V2 languages, for example. Notice that C_{WH} sets WH_D at the moment it enters the derivation c-commanding WH_D (an instantiation Firstness in (14)), and so if C_{WH} is subsequently subMerged by head adjunction of T, WH_D is still indelibly set and ready for assignment when *wh*-Q-phrase is merged to the undominated node.

---

[32]   This will turn out to be a sub-case of scope assignment generally, as in (62).

[33]   McCloskey (2002) argues that 'spurious features' are needed to trigger movement to intermediate positions for long cyclic movement. In Safir, in preparation, I argue that no intermediate triggers are necessary, and that a more elegant account than any other available is feasible precisely when triggers are rejected. It is also unnecessary to introduce an Activation Condition, as in Bošković (2007) to determine what is visible to iMerge, since every node is visible to iMerge.

This is illustrated in the partial derivation in (36), where WH_D is not a real node label, but is the non-terminal (nt) marked with * that contains WH_D, as illustrated in the resulting structure in (36d).

(36)   a. $[C_{WH} [_{WHD} \ldots T \ldots [wh\text{-phrase}] \ldots]]$
              – Merge of C_{WH} results in the setting all it c-commands as WH_D

       b. $[[T C_{WH}] [_{WHD} \ldots T \ldots [wh\text{-phrase}] \ldots]]$
              – T subMerges C_{WH} (as in V2 languages)

       c. $[wh\text{-Q-phrase} [[T C_{WH}] [_{WHD} \ldots T \ldots [wh\text{-Q-phrase}] \ldots]]]$
              – *wh*-Q-phrase first c-commands WH_D (and c-commands C_{WH}), so WH_D is assigned as the scope of *wh*-Q-phrase

       d.

It may not be the case that T-to-C in (36b) is required in every language (even covertly), but I include it to indicate that it does not change how scope is set and then assigned to the *wh*-Q-phrase. The same reasoning applies to the *wh*-Q-phrase in (31), which is assigned its WH_D before it is submerged by late attachment. Discussions of scope assignment to bare *wh*-Q and other quantifiers is reserved for section 7, where what have been called 'covert movements' are addressed.

5.3.   A/A'-Distinction

Reconsider now (28a–b), where *wh*-movement resulted in Case Clash when copies were introduced by iMerge such that two occurrences of the same phrase could be assigned different cases, ruling out garden variety cases of *wh*-movement. Notice that the Case Clash issue for *wh*-movement, or A'-movements more generally, would disappear if the landing sites for A'-movement were invisible to Case assignment by higher heads, that is, once a phrase is A'-moved, it is no longer evaluated for Case mapped onto prominence. Something must insulate A'-moved nominals from further Case assignment or else typical utterances like (28a) would be disallowed.

      As it turns out, VSA does not have to be revised in any way to permit *wh*-movement to avoid Case assignment, at least if a few plausible assumptions about the role of quantificational heads are adopted. For example, it is plausible to assume that *wh*-Q must have scope over its restriction, which consists of the larger phrase that moves with *wh*-Q, just as the *wh*-Q-phrase as a whole must be assigned a scope. Suppose scope over a restriction is achieved by iMerging *wh*-Q to the larger *wh*-Q-phrase it is embedded in where it was first merged. In what follows, I exploit the idea that extraction of *wh*-Q to the margin of the nominal in

which it is embedded insulates that nominal from Case assignment because the *wh*-Q acts as an intervener for Case assignment and is not itself a D that can receive Case by assignment. Consider the following assumptions:

(37)   a.   A *wh*-Q must locally c-command a restriction on *wh*-Q.
       b.   *Wh*-Q is typically embedded in the domain of D.

A tension now arises between (37a) and (37b), since (37b), a consequence of selection, usually, means that *wh*-Q be c-commanded by D while (37a) requires *wh*-Q to c-command D when a whole *wh*-Q-phrase is moved. Given (37a), we expect to see *wh*-Q outside of D in a nominal at the semantic interface, or else the *wh*-Q has not moved to form its restriction (normally resulting in an impossible interpretation). I shall assume that the same condition that applies to interrogatives to create scope assignment also applies to relative clauses, except that the domain is the open sentence formed as a property by iMerge of the *wh*-REL-phrase which is the first to c-command REL$_D$.

Since determiners and quantifiers tend to be mutually exclusive in English nominals, neither claim in (37) can be robustly supported by overt English phenomena. The only overtly moving *wh*-Q-phrases within nominals are instances where there is extraction of the *wh*-Q(-phrase) to the left edge of the pied-piped nominal, as in (38) (see Safir 1986: 679).

(38)   Those reports which, the height of the lettering on, the government prescribes, are tedious.

Cases like (38) are relatively rare in overt syntax for reasons that will become apparent in section 7, but I will assume that extraction of just a quantifier head to form quantifier-restriction structures is the common case in covert (unpronounced) syntax. After all, one appeal of this analysis is that it directly feeds one of the most commonly employed representations of scope, namely, one where the quantifier has scope over its restriction and over the proposition that the restricted phrase originates in (i.e. the nuclear scope; see fn. 33). Moreover, it is an analysis consistent with a proposal independently made by Cable (2008), to which I will return.

Now let us consider the relevant steps in the derivation of the bracketed portion of (39), on the assumption that *wh*-Q is an intervener for Case assignment. The expansion of the tree can be tracked by the brackets on the right, and I pause after (39d) to take stock.

(39)   I wonder [whose mother John likes]
       a.   [[D John] [*v** [like [D [whose mother]]]]]
               – here the EA [D John] has just merged to set its domain as that of *v**
       b.   [[*whose mother*] [D John] [*v** [like [D [*whose mother*]]]]]]
               – under assumptions dating back to Chomsky (1986), the A'-constituent must escape '*v*P' by adjunction (for the reason why in VSA, see section 6).

c.   [[*wh*-Q [D *whose mother*]] [[D John] [v* [like [D [*whose mother*]]]]]]
– *wh*-Q subMerges its restriction and permits the moved phrase to be
invisible for Case prominence

d.   [T [[*wh*-Q [D *whose mother*]] [[D John] [v* [like [D [*whose mother*]]]]]]]
– T assigns Case prominence orders [D *John*] over [D *whose mother*]
assigning R-Case (NOM) to *John* and ACC to [D *whose mother*], while T
agrees with [D *John*](see (39d'))

The key move here is that in (39c), *wh*-Q, just the question quantifier, was
extracted from [*whose mother*] and iMerged to the latter. This is the move that
renders the displaced *wh*-Q-phrase invisible to T, so that the higher occurrence of
whose mother is not ranked for Case prominence with respect to the lower
occurrence of *whose mother*, with the result that Case Clash is avoided. The tree in
(39d') clarifies the Case assignments.

(39)   d'.

When *wh*-Q subMerges [D-ACC [*whose brother*]] before T is merged, a possibility
allowed by Revised Extension, [D-ACC [*whose brother*]], sister to *wh*-Q, is then
insulated from Case Clash. When T merges as in (39d), it orders only [D-NOM
John] >> [D-ACC *whose brother*] for Case prominence, where the latter is sister to
V, and no problem arises. If *wh*-Q does not subMerge its restriction at the first
opportunity after movement, then Case Clash will end the derivation (as soon as
cyclic Spell-Out applies, see section 6) because the occurrences of *whose brother*
have more than one Case assignment between them.

The derivation then con-tinues with further movement of [*wh*-Q [D-ACC
[*whose brother*]]] invisible to Case assignment in higher Case domains.

(39)   e.   [[*D John*] [T [*wh*-Q [D *whose mother*]] [[*D John*]

[*v** [like [*whose mother*]]]]]]]]
– [D *John*] iMerges and satisfies EPP.

f.   [C_WH [[*D John*] [T [*wh*-Q [D *whose mother*]] [[*D John*]

[*v** [like [*whose mother*]]]]]]]]

– Merge of C_{WH} assigns R-Case (NOM) to [D *John*], which can receive it without Case Clash, and C_{WH} sets WH_D

g. [[*wh*-Q [D *whose mother*]] [C_{WH} [[D John] [T [*whose* [D *whose mother*]]
[D *John* [*v** [like [*whose mother*]]]]]]]]]

– the *wh*-phrase reaches its criterial position and is assigned WH_D as its scope

The top of the tree is diagramed in (39g′). If the derivation were to continue beyond the bracketed portion of (39), the Case domain of the matrix T would include the highest occurrence of [*wh*-Q [D *whose mother*]] in its criterial position in (39g), but once again, the *wh*-Q-phrase would be invisible to Case prominence.

(39) g′.

'Pronounce Highest' (see (51) in section 6 below) is the instruction for phonology, but for terminal nodes (that do not exclusively subMerge a terminal), 'Pronounce Lowest' is the order of the day (see (52) below). I will have more to say about how these principles apply in section 6 and, with respect to quantifier scope, in section 7. Thus the highest occurrence of [*wh*-Q [D-Acc *whose mother*]] is the highest non-terminal pronounced, and within that phrase, the second *wh*-Q morphologically embedded in *whose* is the one pronounced as the lowest occurrence of the terminal.

On these assumptions, simpler cases like *whom did John see* require an analysis of *whom* in situ before movement as [D [*wh*-Q *pro*]]. After [D [*wh*-Q *pro*]] moves to its criterial position or to any intermediate one, *wh*-Q will subMerge its restriction, thereby blocking Case from being assigned to D once a Case-assigning head is merged, as in (40).

(40) [H_{Case} … [*wh*-Q [D [*wh*-Q *pro*]]]]

Since D is silent in most languages preceding *wh*-Q and since Pronounce Lowest will apply, [D [*wh*-Q *pro*]] is likely to be heard only in its (highest) criterial position, since it is a higher occurrence of a non-terminal.[34] Notice that if a

---

[34] I am assuming that all lowest occurrences of a quantifier are treated as variables at LF by a rule that converts the copy into a variable (see Safir 2004b and Fox 2003). It is possible that further iMerge of *wh*-Q out of the constituent it forms with its restriction to form a new undominated node, as in (i), to allow the Q to have scope over both its restriction and its domain, but this depends on how the variable-forming rules are formulated.

particular Q is not an intervener for Case, then A'-movement cannot be protected, a consequence that will be explored in section 7.

To summarize, A'-movement of *wh*-Q-phrase is permitted to exist only when the phrase that moves is insulated from Case Clash by the intervention of a *wh*-Q that subMerges its restriction (and for further evidence based on intervention effects, see fn. 43 below). Phrases that are not insulated in this way will be susceptible to Case Clash if they move without R-Case.

Some potential support for this position might be gleaned from a slight reinterpretation of a proposal by Cable (2008), who explores a way to eliminate special mechanisms, like percolation, from playing a role in pied-piping phenomena. He is operating under the assumption that movement to a criterial position is triggered and that the goal probed by C_{WH} should only be a head that takes the 'pied-piped' constituent as its complement. He proposes that a Q-particle he identifies in Tlingit (and, drawing on earlier literature, for Japanese and Sinhala) takes the *wh*-phrase as its complement and it is the maximal projection of the Q-particle that is then moved. He extends this analysis to English, as in (41).

(41) a.   Whose father's cousin's uncle did you meet at the party?
      b.   [**QP** [[[[**whose**] father's] cousin's] uncle] **Q**] did you meet at the party?

He continues (Cable 2008: 22):

> Under this analysis, a pied-piping structure in English is derived exactly like the pied-piping structures of Tlingit. In such sentences, the (null) Q-particle takes as sister a phrase properly containing the wh-word, which entails that the fronted phrase of the wh-question properly contains the wh-word.

The structure of Cable's analysis of the Q-particle is similar to the structure that emerges when the *wh*-Q in a *wh*-phrase is extracted to c-command its restriction, but is pronounced low. Except for the fact that the Q particle in Tlingit has distinct morphology from the wh-phrase (perhaps what permits it to be pronounced high), Cable's analysis suits VSA assumptions very well. The reason for the positioning of the Q-particle may either be an alternative Spell-Out of a *wh*-head, or it may be a head that functions, at least in part, to protect a *wh*-phrase from Case Clash in the course of cyclic movement. The second possibility is explored with respect to another phenomenon in the next section.[35]

---
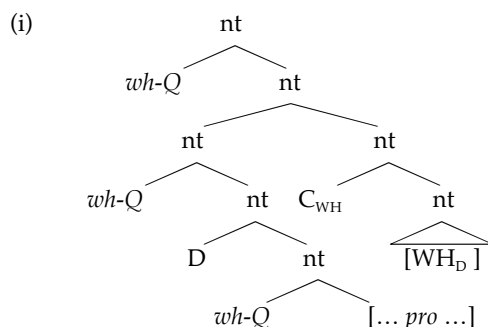
[35]   I will not enter further into Cable's analysis, where complications would surely arise. In particular, he notes that Tlingit pied-piping violates islands and pied-piping in English and

5.4. *Differential Object Marking*

This section closely follows Linares (2008), who demonstrates that VSA permits a much more natural account of differential object marking (DOM) than more conventional accounts based on special Case assignment or marking by a preposition. He proposes that DOM involves the introduction of a head to insulate a shifted direct object from Case Clash.

First, Linares describes the DOM phenomenon succinctly, and I repeat his description here (with examples renumbered to order with mine):

> The phenomenon known as Differential Object Marking (DOM) since Bossong (1985) involves languages in which direct objects appear in two different forms, depending on their intrinsic degree of specificity and/or animacy. In these languages, unspecific and/or inanimate direct objects (DOs) appear in an unmarked, nominative-like form, whereas specific and/or animate DOs are associated to a specialized particle or affix. Thus, for example, specific and animate direct objects in Spanish are associated to the particle *a*, often referred to as 'personal *a*'. By contrast, unspecific or inanimate DOs remain unmarked.
>
> (Linares 2008: 1)

(42) Veo   *(**a**) la amiga   de Pedro.   *Spanish*
      see.1ST   A the friend   of Pedro
      'I see Pedro's (girl)friend.'

(43) Veo   (***a**) una maquina.
      see.1ST   A a   machine
      'I see a machine.'

Linares points out that DOM is widely attested in the world's languages, including Sakha (Turkic), Hindi (Indo-European), Chaha (Semitic), and Miskitu (Misumalpan), many Romance languages, such Catalan and Romanian, as well as Spanish. Then Linares argues for the following generalization, which I name after him.

(44) *Linares' Generalization*

      The marked version of a direct object in a DOM language is (also) a shifted object, while the unmarked version can (but need not) correspond to an object *in situ*.

For example, he shows that specific and/or animate direct objects are marked in Sakha, whereas unspecific and inanimate direct objects are not (46).

---

many other languages does not. He adapts a theory of wh-agreement from Kratzer & Shimoyama (2002) to make the distinction, but it seems that the extra assumptions that are involved in his account, including agreement limited by intervention, are essentially compatible with the Mapping Principle. I leave further exploration of this phenomenon for further work.

(45) a.   Min   Sardaana*(–**ni**)   kördüm.   *Sakha*
           I   Sardaana.ACC   saw.1SG
           'I saw Sardaana.' *(Sardaana: here a personal name, based on a flower name)*

      b.   Ali   bir   piano–**yu**   kiralamak   istiyor.
           Ali   one   piano.ACC   to-rent   wants
           'Ali wants to rent a certain piano.'

(46) a.   Min   sardaana   kördüm.
           I   lily   saw.1SG
           'I saw a lily.' *(sardaana: denotes the flower in this case)*

      b.   Ali   bir   piano   kiralamak   istiyor.
           Ali   one   piano   to-rent   wants
           'Ali wants to rent a (non-specific) piano.'

                                                              (Vinokurova 1998)

Linares shows that unmarked objects must appear in strict preverbal position
(47a–b), whereas ACC objects are placed to the left of a VP-adverb in the un-
marked order (47c).

(47) a.   Masha   türgennik   salamaat   sie–te.
           Masha   quickly   porridge   eat.PAST.3
           'Masha ate porridge quickly.'

      b.   *Masha   salamaat   türgennik   sie–te.
            Masha   porridge   quickly   eat.PAST.3

      c.   Masha   salamaat–y   türgennik   sie–te.
           Masha   porridge.ACC   quickly   eat.PAST.3
           'Masha ate the porridge quickly.'

                                                        (Baker & Vinokurova 2008)

     As Linares points out, however, it is not always easy to know if a
differentially marked object has indeed undergone object shift (OS). Linares
continues (p. 4):

> In Spanish, for example, in which word order is quite free, distributional
> tests fail to diagnose such short movements as OS. Indirect evidence of
> vacuous movement of marked DOs is nonetheless available, in control
> configurations involving gerundival adjuncts. In these contexts, marked
> objects can control PRO in secondary predicates, but unmarked objects fail
> to do so. As a result, [(48a)] is ambiguous, but [(48b)] is not.

(48) a.   Besé$_i$ a una niña$_j$ [PRO$_{i/j}$ llorando].   *Spanish*
           kissed.1ST   A a   girl   crying
           'I kissed a girl while I/she was crying.'

      b.   Besé$_i$ una niña$_j$ [PRO$_{i/*j}$ llorando].
           kissed.1ST   a   girl   crying
           'I kissed a girl while I (*she) was crying.'

                                                              (Torrego 1998)

Further controls support his argument, but I leave these aside. If Linares'
Generalization is correct, and for the sake of argument I will assume that it can be
defended over the enormous range of cases that would have to be tested, then the
question arises as to why such a correlation between OS and DOM should exist.

Linares proposes that the presence of DOM results from the presence of an
intervening head that is inserted to avoid Case Clash when a specific/animate
nominal is moved within the domain of T by OS. For example, the insertion of
the Spanish *a* seems to play the role of insulating head, that is, a head that blocks
Case assignment by T. The difference between DOM and *wh*-movement is that
the movement is within the Case domain of T, assuming movement to a position
that is perhaps higher than v^*, but still to the right of the overt position of the
verb in Spanish. Still, by generating two occurrences within the Case domain of
T, something must insulate the higher occurrence. Linares suggests that the
differential object marker (DOMa) subMerges the shifted object in a derivation
like (49) (showing just the relevant steps, where the external argument is *Joe*) and
the resulting tree is presented in (49f').

(49)   a.   [v^* [see [D Mary]]]

      b.   [[D Joe] [v^* [see [D Mary]]]]

      c.   [[D Mary] [[D Joe] [v^* [see [D Mary]]]]]

      d.   [[**DOMa** [D Mary]] [[D Joe] [v^* [see [D Mary]]]]]

      e.   [T [[[**DOMa** [D Mary]] [[D-NOM Joe] [v^* [see [D-ACC Mary]]]]]]]

      f.   [[D-NOM Joe] [T [[[**DOMa** [D Mary]] [[D-NOM Joe] [v^* [see [D-ACC
              Mary]]]]]]]]

      f'.

Step (49c) is OS. In (49d), [*D Mary*] is subMerged by the DOMa. Notice that (49d)
does not increase the size of the tree, but by subMerging [D *Mary*], DOMa
insulates it from Case assignment by T (by hypothesis). Thus T only ranks the
external argument [D *Joe*] and below it, [D *Mary*], in its lowest, first-merged po-
sition, assigning [D-ACC *Mary*] and [D-NOM *Joe*]. The D of [D *Mary*] embedded in
[DOMa [D *Mary*]] receives no Case, so it does not cause Case Clash with the D-
ACC in direct object position.

Notice that it would not suit VSA architecture for the DOMa to be, itself, a

Case marker, or else Case Clash would result for [D *Mary*] insofar as the higher occurrence would receive two assignments. Linares (2008) argues that the DOMa is not a Case marker, although I will not review his arguments here. However, one may support his view with the contrast between the *a* of the DOMa and the *a* of Romance Dative marking. The Dative *a* argument always corresponds to a clitic in the Dative series, as seen in the double object configuration in (50a) and the causative structure in (50b). By contrast, the clitic corresponding to the DOMa argument is typically from the Accusative clitic paradigm, as in (50c) ((50a–c) from Carlo Linares, p.c.).

(50) a.   Le   dí   una medulla a la   gimnasta / a Michael.
            CL.DAT *gave.1SG a   medal   to the   gymnast.FEM to Michael*
            'I gave the gymnast/Michael a medal.'

      b.   Le   hice limpiar la piscine a Michael.
            CL.DAT *made clean   the pool   to Michael*
            'I made Michael clean the pool.'

      c.   Lo   ví   a Michael.
            CL.ACC *saw.1SG A Michael*
            'I saw Michael.'

It seems that the clitic paradigms can remain true to the Case of the DOMa argument, and this is to be expected if the DOMa is not, itself, a Case assigner or a Case marker.[36]

   This account does not explain why specific and/or animate direct objects should have to undergo OS. A natural way to model OS in VSA would be to introduce a head below T that sets OS_D and then to treat the phrase that first c-commands OS_D as having OS_D as its scope. In other words, OS would be iMerge to a criterial position. It is not at all clear what the semantic value for this head would be, however, beyond just restating the semantic properties of the phrases that undergo OS, but in this respect, the analysis is no more stipulative than most other OS analyses. For OS, however, there is no obvious parallel to quantifier-restriction formation for *wh*-Q to insulate a second occurrence of a nominal under T from Case Clash. In such situations, I have adopted Linares' proposal that first

---

[36]   A systematic aspect of the configuration in (i)-(ii) deserves further mention.

      (i)   [T [$_{CASED}$ … D …]]

      (ii)   [[*D* T] [$_{CASED}$ … *D* …]]

      In (i) and (ii), T is a Case assigning head, so D in $Case_D$ of T will be subject to Case prominence and assignment by T, but D adjoined to T will not be, since it is outside $Case_D$. Now suppose cliticization is subMerge of D to T, and that T protects the D from Case Clash (from above, e.g., from C). For this to work it must be assumed that the clitic is a non-terminal, for example, [D *pro*], as has occasionally been proposed. This would appear to predict the Romance pattern of cliticization, where all of the nominal clitics associated with a verb surface in a clitic row on the highest tensed auxiliary or verb. The next likely target for cliticization would be C, which is also a Case assigner in VSA, and this pattern is also frequently found. Further assumptions are needed to make this work, but the matter deserves more study.

Merge of the DOMa insulates OS nominals from Case Clash.[37]

5.5. *Summary*

The upshot of this discussion is (i) that iMerge is optional in syntax, but (ii) that iMerge typically serves interpretive requirements associated with presence in criterial positions, and (iii) that movement must in general protect multiple occurrences of D-heads from Case Clash. The term A'-movement is now a merely descriptive term for iMerge of constituents that avoid Case Clash in a particular way, that is, by restriction formation, whereby a fronted *wh*-Q head intervenes to block Case assignment from above. OS appears to avoid Case Clash by subMerging a phrasal branch with a special head (DOMa) that also insulates the submerged branch. A-movements typically avoid Case Clash by bearing R-Case or non-structural Case, as discussed in section 4. There is no further syntactic distinction between these iMerge constructions that inherently defines them. No syntactic trigger is appealed to, no stipulated distinction between iMerge and eMerge, and no special appeal is made to 'pair-Merge', as in Chomsky (2004), to account for differences between adjuncts and elements moved to criterial positions.[38]

6.   Spell-Out and Locality

The program so far has been to establish that movements that are treated as triggered by Agree in standard minimalist architecture arise in VSA by free Merge, indirectly driven by interpretation and constrained by Case Clash and other forms of derivational drag. In this section I introduce some of the locality restrictions that regulate long distance movement based on cyclic Spell-Out. A

---

[37]   The distribution of multiple *wh*-interrogatives raise issues for VSA, only some of which get any further mention here. All 'multi-specifier' patterns are problematic for this theory because a domain can only be initially c-commanded once, and so multiple specifiers require further assumptions. Perhaps the second *wh*-phrase tucks in (e.g., as in Richards 1999) under the c-command domain of the first moved *wh*-phrase to permit local c-command of WH_D, or else the second-moved *wh*-phrase subMerges the first moved one, piggy-backing on its domain assignment. Extensions to allow for additional fronted positions are required in every current account, and, except for the VSA advantage with respect to the variety of permitted landing sites under Revised Extension. I leave investigation of these matters for future work, but see Appendix B.
         Moreover, as Jonathan Bobaljik (p.c.) reminds me, domain assignment, as I have formulated it, does not apply to create domains for modal verbs, which can scopally interact with certain QPs. Similar remarks would extend to negation particles in some languages. In the literature on negation, a head+operator analysis has been proposed, for example, Haegeman & Zanuttini (1996) and Haegeman (1995). If I posit a null modal operator that can first c-command the domain M_D set by a modal M, nothing extra has to be said, but I have no evidence for such a claim. An extra statement about quantificational heads may turn out to be necessary, but this I also leave for further work.

[38]   The discussion of criterial positions is incomplete, however, and will remain so, insofar as I have not provided any criterial motivation for scrambling. If scrambling is not criterial, it still must be insulated from Case Clash, and if it is, then it is necessary to posit heads that set scrambling domains. Although both strategies are plausible, I do not pursue the matter further here.

key aspect of the account is that forms that have not been moved soon enough can never be pronounced high, which is another form of derivational drag.

Fox & Pesetsky (2005), developing an idea from Chomsky (2000), suggest that when a certain point is reached in a derivation where cyclic node is merged, every lexical item below that node becomes unalterable with respect to crucial aspects of Spell-Out. Fox & Pesetsky couch this in terms of linearization, which is to say that the linear order of nodes below the relevant cyclic node cannot be altered by subsequent operations. For example, if a cyclic node is merged above A and B such that A is linearized before B (A >> B), and if a subsequent operation were to iMerge B such that a later cyclic node linearizes B >> A, a conflict arises that cannot be resolved, and the sentence crashes in phonology. Overt displacement is then impossible unless the theory allows for lower occurrences not to be pronounced (linearized) in certain contexts. If, however, iMerge precedes linearization, then the lower occurrence can be treated, in some relevant sense, as invisible to linearization. In what follows, I will adopt the essential mechanism just described, but I will not follow Fox & Pesetsky's theory in its specifics.

The version of cyclic Spell-Out developed here distinguishes movement by terminals from movement by non-terminals with respect to how they are treated by Spell-Out under cyclic nodes.

(51)   Only higher occurrences of non-terminals are visible to linearization.

(52)   Only lower occurrences of terminals are visible to linearization.

Only (52) is actually novel here. A node is a lower occurrence if there is a point in the derivation where it is c-commanded by a higher occurrence (unique occurrences are visible according to both (51) and (52)). (52) intrinsically bleeds (51), insofar as (51) will not apply between terminals embedded in non-terminal occurrences, since the terminal in the moved phrase will not c-command its copy in the position of the non-terminal 'trace'. However, if the phrase *which man* is iMerged such that it c-commands its first-merged position, then the first-merged position is a lower occurrence of a non-terminal and not visible to linearization. Further discussion of iMerge and Spell-Out of terminals (heads) is deferred to section 7.

Now, let us assume that the cyclic nodes are C and *v*, which means that when cyclic Spell-Out applies, it applies to everything in the domains of these nodes. However, I do not assume that cyclic Spell-Out applies at the point when C or v is merged to its complement.

(53)   a.   Spell-Out below *v* only occurs when a head selecting *v* is merged.
       b.   Spell-Out below C only occurs when a head selecting C is merged or the derivation ends.

The practical import of this Spell-Out timing is that occurrences (OCCs) below the selectors but above the complement domains of the cyclic nodes (in the *intermediate zone*) are within the purview of (51) and (52). The illustrative indices on

the OCCs in (54) may be thought of as term indices, in the sense of fn. 28.

(54)   nt
         ┌────┴────┐
        H_S   nt
              ┌────┴────┐
            OCC_a   nt
                   ┌────┴────┐
                 OCC_b   nt
                        ┌────┴────┐
                      H_CY   [_nt … OCC_a … OCC_b …]
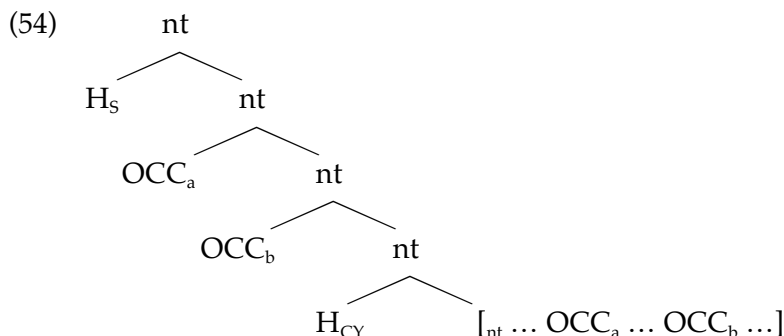
In (54), once the head (H_S) selecting the cyclic head (H_CY) is merged, (51) and (52) can determine which OCCs in the domain of H_CY are to be designated for pronunciation. If OCC_a is terminal, then its lower occurrence is linearized. If it is non-terminal, then the lower occurrence is not linearized. OCCs in the intermediate zone (IZ)—above H_CY but below H_S—are not evaluated for linearization until the next cycle. IMerge will always distinguish higher occurrences from lower ones when a tree is extended in the traditional sense, that is, when iMerge applies to the undominated node (but see the discussion of overt head movement in section 7). As shown in (54), there is nothing in the architecture that limits the number of occurrences that may be in the IZ.

The sort of movement described in (54) will result in well-formed structures (a) if the non-terminal is pronounced high, (b) if it reaches a criterial position that satisfies interpretive requirements, and (c) if it is properly insulated from Case Clash. Consider, however, what would happen if movement of a non-terminal did not stop in the IZ on its way to a criterial position.

(55)   a.   We wonder which boy John saw.
       b.   we [T [v [wonder [*which boy* [C_WH John [T [v [see *which boy*]]]]

In this derivation (irrelevant A-movement suppressed), the OCC of *which boy* in direct object position has been designated for pronunciation because merger of T above *v* has occurred without creation of more than one OCC of *which boy*. However, subsequent iMerge of *which boy* in its criterial position requires that the highest OCC be pronounced, with the result that more than one OCC of *which boy* is pronounced. Using the filter on linearization suggested by Nunes (1999, 2004) and others, locality is enforced by the need to avoid the linearization violation.

(56)   No occurrence can precede itself in phonology.

I do not determine here how cyclic nodes determine what counts as a syntactic island, but limiting access to the IZ will result linearization/realization violations for subsequent extraction of phrases given (51). For example, if no OCC other than the one assigned REL_D is permitted to inhabit the IZ between a relative clause complementizer C_REL (a cyclic node) and the head that selects C_REL

(D or N, depending on other theoretical choices), then extraction of anything but the relative operator from below C_{REL} will result in a linearization violation, because the lowest OCC of any phrase below the cyclic node will have already been designated for pronunciation. The Merge operation is not intrinsically local; locality is enforced by cyclic linearization (and concomitant realization). Such an approach to islands would be compatible with VSA, but it cannot be explored here for reasons of space.

On this account, phrasal movement is always pronounced high, and so there is no space in this theory for covert phrasal movement, that is, phrasal movement pronounced low, as discussed in Appendix B. Covert head movement is discussed at length in section 7.

Finally, a remark is in order about whether or not the assumptions surrounding linearization are consistent with the logic of the KF and VSA. Spell-Out is triggered in this system by the introduction of a head that activates the cyclic node. Since this latter activation is a c-command relation that determines that another c-command-determined domain undergoes an operation outside of syntax (linearization), the only stipulations relevant to KF reasoning are (i) that certain linguistic labels and not others are the ones that set Spell-Out domains, and (ii) that the nodes that select for these labels are the ones that trigger linearization. One might argue that there must be cyclic nodes to limit demand on working memory, for example, but the stipulation as to which nodes are cyclic for Spell-Out is (so far) not derived from the KF and its interaction with the interfaces.[39]

7.   Head Movement and LF Interpretation

Now let us return to the question of head movement, which, according to (52), requires that only its lower occurrence be pronounced. First I will show that head movement to a criterial position, where it iMerges to a non-terminal, need not be cyclic, and therefore is insensitive to cyclic effects like island restrictions. Second, while *wh*-Q is a case intervener, and thus enables A'-movement of the *wh*-Q-phrase, quantifiers that do not block Case will never permit phrasal movement, only head movement, as is the case for most quantifiers. Furthermore, head movement to a criterial position moves without its restriction by definition, and this turns out to be what makes it sensitive to intervention effects, in comparison with overt phrasal movement. Finally, overt head movement is also possible, but only when (52) does not apply. Such cases arise if heads adjoin to heads, such that the iMerged head never c-commands its 'lower' occurrence. Thus this systematic theory of head movement derives many key features that distinguish '*in situ*' quantifiers from the overtly moved sort.

First, consider the fact that the (52) insures that head movement (movement of a terminal) to a position where it c-commands its copy will always require that its lower copy be the pronounced copy. Now consider what would

---

[39]   Bošković (2007) contends that all phrasal nodes are cyclic, and if so, a version of VSA empirically compatible with this broader application would come closer to satisfying the burden of the KF, but I have not explored this possibility for reasons of space.

happen if *wh*-Q were to move just as phrases do, first to an intermediate position, and then to a criterial position, as in (57).

(57)   a.   [T … [*wh*-Q [EA [*v** [V [D … *wh*-Q …]]]]]]
       b.   [*wh*-Q [C$_{WH}$ [… T … [*wh*-Q [EA [*v** [V [D … *wh*-Q …]]]]]]]]]

Under at least one interpretation of 'Pronounce lowest occurrence', head movement will always be restricted to swoop movement in this account, that is, a single movement from first merge position to its criterial position. If head movement (to a c-commanding position) were cyclic, then even after the first merged position is linearized, higher occurrences of the head would still be lower than subsequent head movements, with the result that 'Pronounce lowest occurrence' would pronounce the intermediate head, with a resulting linearization violation with respect to the lowest occurrence. Thus the lowest terminal node copy will be spelled out in the first linearized cycle that contains it. But the empirical results of the theory of cyclicity in the last section, i.e. all the locality restrictions, arise because Spell-Out must be avoided when it is too low. The game is over before it begins for lowest terminals, given (52). Nothing prevents a terminal from moving any distance at all as long as it ends in a criterial position. Thus the theory makes the prediction in (58):

(58)   Movement of quantifier heads is
       a.   pronounced low (*in situ*),
       b.   insensitive to islands, and
       c.   moves without intermediate stops.

These results are largely consistent with what has been observed about *wh-in-situ* languages (e.g., Huang 1982 on island effects), although the usual puzzle concerning the clause-boundedness of most non-*wh* quantifiers remains underived and unresolved (as it is for most current theories), apart from assumptions about Beghelli & Stowell's (1997) clausal architecture discussed in the next section. Whether or not (58c) is correct would depend on showing that *in situ* quantifiers never have intermediate scope, a prediction that depends on too many other assumptions to be examined further here.

     Now recall that I have also assumed that *wh*-Q acts as an intervener for Case assignment. This intervention prevents Case assignment to *wh*-phrases in intermediate positions, thereby avoiding Case Clash, thanks to restriction formation involving subMerge on the left branch. This configuration was illustrated in (40), repeated here.

(40)   [H$_{Case}$ … [*wh*-Q [D [*wh*-Q *pro*]]]]

Suppose, however, that only *wh*-Q is an intervener for Case and other Q heads are not. In this system, there are three immediate consequences for Q heads that are not Case-interveners (call them 'Q$_O$'). First, restriction formation in a left branch will not prevent Case assignment from applying across Q$_O$ to D, creating a Case Clash if there is more than one occurrence containing D in Case$_D$. This

entails that phrasal movement to a criterial position for such quantified nominals is normally impossible, since the D in the higher copy (bolded) will get a different Case assignment from that of the lower copy.

(59)   *$[H_{Case}$ … [[Q_O [D [Q_O *pro*]]]] H_o [… [D [Q_O *pro*] …]]]

Thus all Q_O must find their criterial positions by head movement, never by phrasal movement, and be pronounced low by (52). These are the cases known in the literature as 'covert movement' and they will be further discussed in section 7.2.

The immediate empirical consequence for quantifiers that are pronounced low is that their restrictions are never high. As a result, Principle C effects are predicted in cases like (60), where the quantifiers in the Beghelli & Stowell classification (see section 7.1) would find criterial positions above *v*P by head movement, but direct objects would still c-command the restrictions on the quantifiers (e.g., *books that criticized Noonan* and *book about Mary* in (60a–b), respectively).

(60)   a.   * Sheelagh gave *him* [many books that criticized *Noonan*].
       b.   * Arthur sent *her* [every book about *Mary*].
       c.   *Richard finally told *them* about three critiques of *the teachers*.
       d.   Which critique of *the teachers* did Richard finally tell *them* about?

Judgements about the success of coreference in (60d) are not uniform, but even those who reject the overt phrasal movement example in (60d) typically report it as less deviant than the *in situ* cases in (60a–c) (see Safir 1999). The difference is that overt movement has moved its restriction out of the c-command domain of the pronoun, and if the copied restriction is interpreted high, then it will not induce c-command effects. For (60a–c), which in this theory are predicted to be derived by head movement (since the quantifier is pronounced low), the restriction never moves with the Q_O, hence it can only be interpreted low, predicting the Principle C effects.

Now consider contexts where heads appear to have overtly moved, as, for example, in V2 constructions and in constructions involving head-to-head movement such as T to C. We have just established that head movement to the undominated node in this system cannot be overt because of (52), which will insure that the lower occurrence of the head will always be favored for pronunciation. However, it is possible for head movement to be overt if there is no way to determine which of two occurrences is higher. This situation will arise when head movement proceeds by submerging a head without submerging its sister, that is, in contexts of head-to-head movement.

(61)   a.   [H_1 [… H_2 …]]
       b.   [[H_2 H_1] [… H_2 …]]

Although head movement resulting in (61b) is the most widely assumed analysis of the overt displacement of heads, only Revised Extension permits such

structures to arise from (61a). In (61a), H_1 is higher than H_2 because H_1 c-commands H_2, but in (61b), neither occurrence of H_2 is higher than the other because there is no c-command relation between the two occurrences at any point in the derivation. When this occurs, the system does not predict which node is pronounced, but something must insure that only one of the nodes is linearized. Although I will treat cases like (61b) as the source of overt head movement structures, I regard such cases as outside the core of the linearization theory (embodied in (51-2)), and hence a matter for particular morphologies, with a certain amount of linguistic variation as a result.[40]

7.1. Domain-Setting for Scope

One of the major contentions built into VSA is that all prominence relations are keyed to domain-setting and domain assignment in the course of a derivation and this has been illustrated for Case, argument structure and agreement. Domain assignment for *wh*-Q-phrases has also been instantiated as first c-command of a domain set by C_{WH} to license the interpretation of a *wh*-Q-phrase. Scope interaction can be established by the same mechanisms.

Suppose, for example, that the domain Q_D for a certain quantifier Q is set by a functional head specific to that (class of) quantifier, as in Beghelli & Stowell (1997). The domain of a quantifier is assigned in the same way that domain assignment applies generally, i.e. like external argument assignment, and assignment of *wh*-scope as in (34) is a sub-case of (62)).

(62)   *Assigning Scope*

The scope of a quantifier Q is assigned when Q is assigned a domain that matches its quantificational features / properties.

---

[40]   I am avoiding discussion of the voluminous literature on head-to-head movement here. I concur with Matushansky (2006) (a useful re–assessment of the issues) that the movement is not phonological, but I do not regard it as different in nature from phrasal movement insofar as both can involve subMerge, given Revised Extension. Thus I do not appeal to a special morphological merger rule that must apply to deform the tree after normal iMerge, as she proposes. Nothing in my theory predicts that head-to-head movement must be local, although linking it to intervention for c-selection, as Matushansky does, would suit VSA for movement of heads to positions that are not criterial, yet c-command issues remain. Like Matushansky, I abstract away from some aspects of word internal structure such that what I am calling a head is the node that bears a label under which further morphological analysis is hidden from Merge in syntax, but I have no principled account of that divide. Most distinctly from Matushansky, I do not distinguish head vs. phrasal movements by their triggers, since there are no operational triggers in VSA. It could be that there is morphological longing for affixation that head-to-head movement satisfies or fails to for particular morphologies, but this again leads us into assumptions about particular morphologies and the phonological nature of affixes.

With respect to situations where more than one copy is pronounced, Nunes (2004) appeals to morphological reanalysis in cases where atypical Spell-Out results, as in certain cases of copy doubling. It is also possible that the lack of c-command between copies plays a role in conditioning these atypical outcomes. Sideward movement of any kind would be ruled out by (56) if there is no way to reduce pronounced occurrences to just one where (51–52) do not apply.

Note that on this account, a quantifier can only be ranked for scope after it has moved to a scopal position (not *in situ*), a point to which I will return. Relative scope arises when the domain of one quantifier is in the domain of another, as in (63), where $Q_1 \gg Q_2$ because the domain of Q_2 is inside the domain of Q_1.

(63) $[Q_1 \ [H_1 \ [Q_{D1} \ ... \ Q_2 \ [H_2 \ [Q_{D2} \ ...]]]]]$

The quantifier last to reach its criterial scopal position will thus have widest scope.

Notice that iMerge of Q_1 cannot fall short or bypass the position where its scope domain is assigned, or interpretation fails. Similarly, interpretation fails if Q_1 or Q_2 is assigned a domain set by a head that is inappropriate for it (i.e. this could be thought of as the VSA analog of an interpretable feature that is not interpreted). For example, if H is interrogative C_{WH} which sets WH_D, then iMerge of *every* to $[C_{WH} \ WH_D]$ will mean that *every* is assigned WH_D. An interpretive mismatch results, since *every* is not interrogative.

(64) #$[every \ [C_{WH} \ [_{WH_D} \ ... \ every \ man \ ...]]]$

The mismatch in question is only at the level of interpretation, however; the syntactic structure in (64) is well-formed.

It must be noted, however, that the account offered so far does not insure that a Case-intervening Q (like English *wh*-Q) will achieve its criterial position by overt phrasal movement—head movement by *wh*-Q should also be possible. For languages like English, where at least one overt wh-phrase must be iMerged, an additional stipulation must be made, namely, (65).

(65) WH_D must be assigned to a non-terminal.

(65) resembles the EPP in (20) (and is no more explanatory), but while the EPP may be universal (accounts differ), it is clear that (65) is not.[41] As long as (65) is satisfied, subsequent *wh*-movement, as in multiple interrogations in English, can be by head movement. Unlike iMerge trigger theories, however, this approach is consistent with Rizzi's (1990: 46–48, 2006) suggestion that *why* is generated directly in its criterial position, without the application of Agree to activate an unvalued feature in its domain, consistent with his approach to criterial assignment in Rizzi (2006). With respect to other language specific stipulations, see Appendix B.

7.2.   *Scope Ambiguity*

As mentioned, VSA is consistent with clausal architectures that generate heads associated with scopal positions, as in Beghelli & Stowell (1997) (henceforth, B&S) and references cited there. B&S provide a typology of quantifiers and they

---

[41]   Jonathan Bobaljik (p.c.) points out that first position in V2 languages may require a language-class specific EPP-like statement as well.

suggest that the scopal positions that a quantifier can be interpreted in are determined by the type of quantifier it is. In their way of putting things, the quantifier Q must be found in the specifier of the head that provides a domain Q is licensed to have scope over. This model lends itself to VSA architecture, in which a head sets the domain and the first c-commanding phrase merged above it is assigned the domain-scope by (62).

Prominence for scope will arise in VSA when the domain of one quantifier contains the domain of another one. If neither of two scope-interacting Q heads are interveners for Case, then both will be assigned scope after head movement to a criterial position and both will be pronounced low. However, in the B&S architecture, some quantifiers have more than one criterial position, and when a given quantifier can move to a criterial position either above or below some second quantifier, then scopal ambiguity will arise, since the derivational outputs will allow two different structures for a string. The output of phonology will not distinguish the two structures in this respect, though their structures lead to distinct interpretations. This is the source of scope ambiguity in VSA.

To get a practical sense of how VSA represents relative prominence for scope, consider the two scope possibilities permitted by (66).

(66)   Two soldiers praised every general.

When *two soldiers* is merged to *praise every general* to form *two soldiers praise every general*, *two soldiers* is the more prominent argument of *praise* because *two soldiers* is assigned external argument status when it first c-commands the selected domain of v^*, v^*_D, and v^*_D contains *every general*. This argument prominence relation, that is, *two soldiers >> every general* for *praise* is thus fixed before any further operation, and no subsequent Merge operation can reverse any prominence relation once it is set. If the scope-domain-setting (SDS) head H_1 of the right type is merged to the constituent containing the full argument structure for *praise*, then for the quantifier *every* to receive the proper scope assignment, iMerge must apply to the terminal *every* to yield (67a). This movement is optional in syntax, but if it does not occur then the quantifier *every* will not be assigned a proper scopal interpretation.

(67)   a.   [*every* [H_1 [two soldiers [v^* [praise *every* general]]]]]
        b.   [*two* [H_2 … [*every* [H_1 [*two soldiers* [v^* [*praise every general*]]]]]]]]

When H_1 is merged it sets the contents of its sister, H_{1D}, as its domain, and when *every* is iMerged so that it first c-commands H_{1D}, *every* is assigned H_{1D}, as it is in (67a). This means that *every* has scope over every other quantifier assigned a domain that is contained in H_{1D} (up to intervention), but in (67a), no such assignment has been made (*two* has not been assigned a scope yet). Now *two* must be assigned a scope if it is to be interpretable. Just as with *every* of *every general* and for quantifiers in general, *two* of *two soldiers* can receive an interpretation only if *two* can be assigned the right sort of domain. So if the right sort of scope-domain setting head H_2 is merged above the higher occurrence of *every*. Then iMerge of *two* will assign to it H_{2D}, the domain set by H_2. Since H_{2D}

contains *every* and the domain that has been assigned to it, scope is established as *two >> every*.

Once scope prominence has been assigned in this way it cannot be reversed later in the derivation, or Scope conflict arises, and no scopal interpretation is possible (i.e. another form of incoherence). In other words, relative scope assignment becomes a form of derivational drag like Case Clash and linearization paradoxes. If *two* had moved to a criterial position first, and *every* after, then the prominence relation would have been reversed.[42]

What has been schematically presented in (67) would be instantiated in the B&S system where the head corresponding to H_1 is Dist (the head just below T that sets the domain for 'distributed' quantifiers like *every*) and the head corresponding to H_2 is Ref (one possible landing site for 'group' quantifiers, like *two soldiers*). The interpretation for (66) with wide scope for the universal would be achieved by merging a SDS head, Share (= H_1), above v^*, where Share would set its domain (Share$_D$). When *two* is iMerged above Share, as in (68a), *two* first c-commands Share$_D$, thereby assigning the scope domain for *two* as Share$_D$.

(68) a.   [*two* [Share [*two soldiers* [*v** [*praise every general*]]]]]
      b.   [*every* [Dist [*two* [Share [*two soldiers* [*v** [*praise every general*]]]]]]]

Every gets its scope in the same manner that it did in (67a), that is, Dist is merged to the constituent in (67a), such that Dist$_D$ is set, and then Dist$_D$ is the scope domain assigned to *every*. Since Dist >> Share (universally, in the B&S theory), *every* has scope over *two*.

The key to the ambiguity of scope in the B&S analysis is that group quantifiers have more than one landing site because they can generally be associated with more than one scope domain, whereas distributed quantifiers are not so flexible. The derivations in (67) and (68) assume that the first quantified phrase that first c-commands a domain set by the head appropriate to it, will end up having the narrower scope. Extension will conspire with iMerge to insure that any other quantified phrase will have to be Merged to a criterial position above the first one. Any theory that distinguishes quantifiers in this way is suitable for instantiation in VSA, and so I will not further explore the particulars of the B&S system.

7.3.   *Scope and Intervention*

In other sections I have discussed a variety of intervention effects where domain-setting heads are interveners for other domain setting heads of the same kind: Case assigning heads and Q are interveners for other Case assigning heads, selecting heads for other selectors, agreeing heads of one sort or another are

---

[42]   Actually, B&S assume that 'group' quantifiers, of which *two soldiers* is an exemplar, can be interpreted in situ. If so, then there is an interpretation for (67a) for which *every general* has scope over *two soldiers* without movement of *two*. I have nothing to say about in situ interpretation of certain classes of quantifiers, but it is less natural in my system. I will assume for the sake of presentation that all quantifiers must establish scope by means of a domain set by a compatible head.

interveners for heads of that sort, and cyclic heads intervene to render their domains opaque to linearization by higher cyclic heads. This section explores intervention effects that arise when certain *in situ* quantifiers are in the domain of certain intervening quantificational heads. Given the approach to the overt/ covert displacement developed here, scope intervention effects typically arise where the lowest occurrence of a scope-bearing head is separated from its highest occurrence by an intervener (typically a certain class of SDS head). Beck & Kim (1997: 370) state the relevant effect as follows:

(69) a. *Quantifier Induced Barrier*

The first node that dominates a quantifier, its restriction, and its nuclear scope is a Quantifier Induced Barrier.

b. *Minimal Quantified Structure Constraint*

If an LF trace β is dominated by a quantifier induced barrier α, then the binder of β must also be dominated by α.

Following (and interpreting) the literature, it will be argued that pied-piping is a means of 'smuggling' a lowest Q-head occurrence past an intervener, thereby neutralizing the intervention effect, given a plausible treatment of the relation between copy theory and variable formation (see fn. 33 above).

Pied-piped constituents, since they are phrasal, will always be pronounced high in VSA, so this theory is designed to predict that intervention effects are only found in cases of head movement to a criterial position, where the lower occurrence is pronounced, as is generally the case for intervention effects (but see (77-78)).

Consider the following intervention effects that have been noted for Korean by Beck & Kim (1997) for (70)–(72) and by Kim (forthcoming) for (73)–(74).

(70) a. * Amuto   muô–s–ûl   ilk-chi   anh–ass–ni?   *Korean*
           *anyone   what.*ACC   *read-*CHI *not-do-*PST*-*Q

      b. Muô–s–ûl   amuto   __   ilk–chi   anh–ass–ni?
          *what-*ACC   *anyone*   *read-*CHI *not-do-*PST*-*Q
          'What did no one read?'

(71) a. *Minsu–man   nuku–lûl   po–ass–ni?
          *Minsu-only   who-*ACC   *see-*PST*-*Q
      b. Nuku–lûl   Minsu–man   __   po–ass–ni?
          *who-*ACC   *Minsu-only*   *see-*PST*-*Q
          'Who did only Minsu see?'

(72) a. *Minsu–to   nuku–lûl   po–ass–ni?
          *Minsu-also   who-*ACC   *see-*PST*-*Q
      b. Nuku–lûl   Minsu–to   __   po–ass–ni?
          *who-*ACC   *Minsu-also*   *see-*PST*-*Q
          'Who did Minsu, too, see?'

(73) a. ?*Nwukwun–ka–ka   mwues–ul ilk–ess–ni?
          who-some-NOM   *what-ACC*   *read-PST-Q*

      b. Mwues–ul nwukwun–ka–ka   __   ilk–ess–ni?
          what-ACC   *who-some-NOM*   *read-PST-Q*
          'What did someone read?'

(74) a. *JOHN–I   mwues–ul sa–ss–ni?
          John-NOM   *what-ACC*   *buy-PST-Q*

      b. Mwues–ul   JOHN–I   sa–ss–ni?
          what-ACC   *John-NOM* __ *buy-PST-Q*
          'What did JOHN (but not someone else) buy?'

In all of these cases, the a-examples show that an intervener blocks wide scope
interpretation for the in situ quantifier and the b-examples show that wide scope
is available for the same quantifier if it moves overtly to a position higher than
the intervener. Negation intervenes for an in situ question in (70a), but scrambl-
ing of the *wh*-phrase to initial position in (70b) is acceptable. The interveners for
(71)–(74) are 'only', 'also', existential-indefinite *-ka* marking, and contrastive
focus, respectively, and in each case, if the *wh*-phrase is scrambled to initial po-
sition, as in all the (b) examples, the intervention effect disappears. The contrast
is schematically modeled for (70a–b) in VSA in (75a–b) where Neg (underlined) is
the intervener for Q.

(75) a. * [Q [C_Q [_Dq ... [[Neg-XP] [Neg [_Dneg ...[D [Q N]] ...]]]]]]
      b.   [[Q [D [Q N]]] [C_Q [_Dq ... [[Neg-XP] [Neg [_Dneg ... [D [Q N]] ...]]]]]]]

Notice that the negative quantifier phrase is assigned scope in the usual way
according to VSA, that is, an SDS head (Neg in (75a)) has set a domain, Neg_D, that
has been assigned to the quantifier or quantified phrase (for a clausal architecture
for Korean compatible with these assumptions, see Kim, forthcoming).

        Three points are of particular significance here with respect to the contrast
between the a- and b-examples in (70)–(74), respectively. First, the a-examples in-
volve direct head movement to a criterial position, leaving a lower occurrence in
the form of a head, while the b-examples involve overt movement of a phrase
(with head movement within the phrase of the Case-clash-insulating head Q).
'Smuggling' movement in (75b) does not have to be directly to a criterial position,
as long as it is outside of Neg_D. Second, the rest of the quantified phrase in the a-
examples, the part that would count as its restriction if the quantifier were
moved, does not have an occurrence above the intervener. Third, the intervener,
underlined in (75a–b), is an intervening SDS head that has set a domain for some
other quantifier; the quantifier associated with the intervener is not the actual
intervener in this system. Interpreting proposals in the literature (e.g., Pesetsky
2000: 67) in terms of VSA, I describe the effect as follows:

(76) The occurrence of Q in its criterial position cannot be separated from at
      least one occurrence of its restriction by an intervening SDS head of type Y.

The reason, then, that the b-examples in (70)–(74) permit wide scope is because phrasal movement has moved Q's restriction higher than the intervening SDS head, so interpretation has access to the restriction of Q.[43]

It is important to keep in mind that not all intervention effects involve the failure of wide scope readings for *in situ* quantifiers.[44] Overt movement that strands part of the restriction of a quantifier also falls under the generalization in (76). Indeed the first known effects of this kind, such as that reported for French by Obenauer (1984) in (77) and that reported for German by Beck (1996) in (78) (from Pesetsky 2000: 68) involve overt movement of a portion of the quantified phrase that may or may not be an instance of movement by a head.

(77) a.   Combien   de livres a–t–il beaucoup consultés __?
            how-many   of books did-he a-lot   consult
            'How many books did he consult a lot?'

      b.   Combien   a–t–il (*beaucoup)   consultés [__ de livres]?
            how-many   did-he a-lot   consult   of   books
            'How many books did he consult a lot?'

                                                                                    (Obenauer 1984)

(78) a.   [Wen von dem Musikern] hat   keine Studentin __ getroffen?
            whom of   the musicians has   no   student   met
            'Who among the musicians has no student met?'

      b. * Wen hat   keine Studentin [__ von dem Musikern] getroffen?
            whom has   no   student   of   the   musicians met
            'Who among the musicians has no student met?'

                                                                    (Beck 1996, Pesetsky 2000: 68)

The movements in (77b) and (78b) are exceptions to (52) if they are head movements pronounced high, but they are consistent with (51) and (52) if they are partial phrasal movements with part of the moved phrase silent. Even in the latter case, (76) applies to rule out (77a) and (78a), but not (77b) and (78b).

Much of what I have said so far is a translation of scope intervention into VSA, but the correlation proposed here between quantifier head movement, pronunciation low (*in situ*) and the potential absence of island effects makes a prediction within VSA that is distinct from theories that rely only on the *Agree*

---

[43]   I have identified the SDS head as the intervener rather than the domain of that head or the quantifier that is assigned that domain because the SDS head, as a terminal node, has a label, but I have no evidence that slightly more complex statements in terms of intervening quantifiers (or quantified phrases) or containing domains (perhaps derived as a semantic consequence) could not achieve the same result. The application of (76) depends on choices for Q and Y, and in the absence of a way to predict which interveners block which quantifiers (i.e. no strong adherence to the Natural Intervention Hypothesis), I will not further explore this matter here (but see Appendix B).

[44]   Sauerland & Heck (2003) show that there are intervention effects internal to pied-piped phrases when what I am calling restriction formation needs to take place. They show that overt movement within the pied-piped phrase or the avoidance of pied-piping altogether, results in well-formedness. In other words, intervention effects place a limitation on the embedding of the scope-bearing head in pied-piped phrases. See their paper for details.

relation to predict locality. Even when an *in situ* quantifier can be construed outside an island in the absence of an intervener, the presence of an intervener outside of an island is still predicted to cause an intervention effect. Consider the Korean examples in (79) (provided by Hyunjoo Kim, p.c.).

(79) a.　　Mina–ka　　[[nwukwu–lul　al–nun]　　salam]–ul　po–ass–ni?
　　　　　　Mina-NOM　who-ACC　　　know-REL　person-ACC　see-PST-Q
　　　　　　'Mina saw the man who knows who?'

　　　b.　*??Mina–man　　　[[nwukwu–lul　al–nun]　　salam]–ul　po–ass–ni?
　　　　　　Mina-ONLY　who-ACC　　　know-REL　person-ACC　see-PST-Q
　　　　　　'Only Mina saw the man who knows who?'

　　　c.　*??MINA–KA [[nwukwu–lul　al–nun]　　salam]–ul　po–ass–ni?
　　　　　　MINA-KA　who-ACC　　　know-REL　person-ACC　see-PST-Q
　　　　　　'Mina (and no one else) saw the man who knows who?'

The *in-situ wh*-quantifier in (79a) is embedded in a relative clause (marked by brackets), from which overt movement is not possible, yet (79a) is successfully construed as a non-echo direct question. In (79b), however, the presence of the intervener hosting *Mina-man* outside of the island, but below the high scope position required for a matrix interrogative, results in an intervention effect. In (79c), it is the contrastive subject that causes intervention. The persistence of the quantificational intervention effect where cyclic movement is not possible suggests that *intervention applies over a potentially unbounded domain*, that is, no matter how large a span a single iMerge head movement can cover, intervention will hold across that potentially unbounded distance. If Agree were responsible for the quantificational intervention effect in (79), then the assumption that Agree is bounded by a cyclic domain must be abandoned. However, such a move is tantamount to conceding that an interpretive relation R stated on a c-command relation is the basic sort of viable relation, and that bounded c-command relations, including those that involve Case and agreement, simply involve more local interventions in most instances. Further inadequacies of Agree as an alternative to c-command are discussed in Appendix A.

7.4.　Summary

The approach to head movement developed in this section aligns it with what has been called covert movement or LF-movement in the literature, but not with covert *phrasal* movement. In VSA, head movement to a c-commanding position is always pronounced low and typically is movement to a criterial position establishing scope. Overt head movement arises when one head subMerges another head, an outcome that is not regulated by (52), since the iMerged occurrence does not c-command the 'lower' one. Head movement to a c-commanding position is unbounded movement insensitive to islands, as opposed to overt movement, because islands, insofar as they are understood, are dealt with by cyclic Spell-Out, not by quantifier intervention. The inapplicability of cyclicity to head movement is principled, because cyclicity only leads to violations that arise from lineari-

zation conflicts, and movement uniformly pronounced low will never cause linearization conflicts.

Head movement to a criterial position strands its restriction, which is then unavoidably susceptible to Principle C effects and intervention effects, superiority perhaps being one of the latter. Intervention effects are neutralized by overt movement above the intervener because overt movement is necessarily phrasal (by (51)). Phrasal movement to a criterial position is typically a form of pied-piping in this system, and it is always an option if Case Conflict or interpretation of quantifiers or linearization restrictions do not rule it out. An EPP-like feature in C can have the effect of forcing movement of a non-terminal, as in English questions, but how large a non-terminal moves is not determined by any device. The latter result leaves 'heavy pied-piping', such as pied-piping of more than a nominal, unexplained, as it is in other accounts—but unlike Spec-Head feature checking theories, no ad hoc feature-percolation is needed to justify it (movement is always optional) and indeed, no Spec-Head checking can even be formulated in VSA.

8.   Conclusion

If VSA can be feasibly maintained, then the wide range of unselected-for structural complexity that occurs in natural language can be traced to interaction of two simple factors, pre-HLF capacities, including Merge, and the Mapping Principle. The case for the existence of a KF is thus more plausible, and even if the Mapping Principle proves insufficient, it has been part of the goal of this essay to demonstrate how high the bar must be set if we are to meet the burden of the KF.

I have argued, however, that Merge and interface interpretive relations mapped on to c-command, interacting with the properties of lexical items, inherent prominence relations, and relative closeness, are indeed sufficient to generate the complexity we observe in natural language syntax. Attempts to streamline the theory so that only these notions were appealed to required solutions to prominence conflicts without appealing to operational distinctions between iMerge and eMerge, or labels projected to non-terminal nodes, or Spec-Head feature-checking or numerations.

Instead, I have featured the role of c-command as the template for all mapping relations interpreted at the interfaces. Heads set domains by c-commanding into them up to intervention, sometimes by necessarily selecting heads in domains (c-selection), sometimes by interacting with other heads in their domain if there are any (agreement), sometimes by prominence ordering of elements in their domains (Case prominence), and sometimes by setting the domains necessary for criterial assignments, as in the case of scope assignment. The special role of unique specifiers in other theories has essentially been derived by first c-command (as in domain assignment for scope) which is a continual source of derivational drag. A further role for c-command is post-derivational, as illustrated by the Korean island-insensitive intervention effects, which also demonstrate the insufficiency of Agree as an alternative addition to Merge. Further

evidence that Agree cannot replace appeal to c-command for the statement of structurally sensitive interpretive relations is provided in Appendix A.

As VSA emerged, I did appeal to special notions regulating Case (for example, (18)) and Spell-Out (for example, (51-52)) that do not follow from Merge and c-command mapping, but that do have to do with the way in which the output of Merge is evaluated at the phonological interface. The theory of interpretation in criterial positions does not appear to need anything novel in this account, except revising Extension to allow instances of subMerge on an immediate daughter of the undominated node, a revision that is independently necessary for overt head movement, tucking in, and late attachment. Failures of (appropriate) scope assignment also result in ill-formedness, where failure of interpretations is a filter on derivations.

Most of the component parts of VSA have been adapted from existing proposals and analyses, and thus rely heavily on generalizations and mechanisms that others have explored, including the prominence theory of Case assignment (particularly Marantz 1991 and Bobaljik 2006), the theory of criterial positions (particularly Rizzi 2006), the theory of cyclic linearization (particularly Chomsky 2000 and Fox & Pesetsky 2005), the nature of scope assignment (Beghelli & Stowell 1996) intervention (particularly Beck & Kim 1997 and Pesetsky 2000). The Mapping Principle, however, is what makes sense of why these components have the form that they do and how they interact. The interactions have produced some novel results, among them the dissolution of any theoretical A/A'-distinction, which is now only an artifact of the strategy by which Case Clash is avoided, and an explanation of why the scope of *in situ* quantifiers is acyclic, unless intervention effects on heads or restrictions keep it local. All of these components and results have been sewn together as manifestations of the role of the Mapping Principle in the structure of linguistic architecture.

Many issues not touched on here will have to be explored if VSA is to succeed as a leading idea informing the relationship between evolutionary boundary conditions on the emergence of human cognition, on the one hand, and the rich world of linguistic structure and diversity, on the other. This essay is only the first step.

Appendix A: C-Command vs. Operational Agree

In recent work, it has been suggested that c-command may not need to be independently stated in linguistic theory because all of the cases for which it is required can be handled by operational Agree (e.g., Chomsky 2007a, 2008). Note that Agree is assumed in these works to be independently necessary in addition to Merge, and thus positing Agree is, for those who believe that Merge is the KF, a respect in which the burden of the KF is not met. The only role of this section is to show that the very few arguments put forth as evidence for this view, all based on anaphora, do not go through unless (i) operational Agree relations are multiplied for local relations and (ii) an additional device is introduced as a c-

command based interpretive relation to apply over long distances.

With respect to relations that appeal to unbounded c-command, Boeckx (2003) employs 'Match' from Chomsky (2000: 122),[45] a device similar to many in the literature in this respect, to account for the binding of resumptive pronouns in weak islands, along with Agree, and also 'intrusion', which permits resumption into strong islands (and must be c-command sensitive), as in a typical Irish example from McCloskey (1979: 34).

(A1) Sin   teanga   aN mbeadh meas agam   ar duinear bith   aL tá
       that a-language aN would   be   respect at me   on person aL is
       ábalta   í   a labhairt.
       able   it to speak
       'That's a language that I would respect anyone that could speak it.'

(McCloskey 1979: 34)

Moreover, intervention effects must have the same open-ended character to cross islands as we have already seen with respect to the Korean cases in (79), where intervention effects hold even if the moved element could not have moved to a position locally below the intervener (in the same cycle), so the intervention must hold over distances greater than those permitted by Agree. Moreover, many languages employ a logophoric form of a pronoun which must be used when it is anteceded by the reported speaker in the matrix clause. This is true even when the pronoun is embedded in an island. Clements (1975: 156) reports that a relative clause complement to the verb meaning 'remember' in Ewe cannot embed the logophoric form referring to the one who remembers (*yè* is ill-formed in place of *e* in (A2a)). However, *yè* becomes possible if the whole structure including the relative clause is embedded in the CP complement to a verb of saying as in (A2b).

(A2) a.   Ama do nku nyɔnuvi hi dze e   gbɔ dyi.
             Ama set eye girl   wh stay 3.PS side on
       b.   Ama gblɔ be   yè-donku nyɔnuvi hi dze yè gbɔ dyi.
             Ama say that yè set eye girl   wh stay yè side on
             '*Ama* said that *she* remembered the girl who stayed with *her*.'

(Clements 1975)

These arguments from resumption, intervention and logophoricity provide a wealth of evidence that unbounded c-command still plays an important role at the interpretive level.

Nonetheless, Chomsky (2007a, 2008), Reuland (2005), and Kratzer (2009) have suggested that appeal to Agree could displace the need for c-command. They argue that the locality principle for anaphora known as Principle A follows from Agree, either because Agree can do all that is required (e.g., Kratzer) or that

---

[45]   Match is supposed to involve matching features, but as Seely (2006: 202) points out, selection cannot be Agree, but neither is it Match, since the selected feature does not match the categorial feature of the probe in any obvious way.

Agree can achieve binding relations that standard c-command cannot (Chomsky 2007a, 2008, following Reuland 2005).

In VSA, Case and agreement relations are established by heads that establish c-command domains bounded by interventions. In section 4.3, the distinctions between Case and agreement relations for T meant that T was in a local relation of one sort for agreement, and of another sort for Case. In other words, not all local relations are Agree relations. The same sort of argument can be made with respect to the separable locality relations that must be appealed to account for morphological verb agreement and anaphora relations. Since Agree is a local relation and it is a relation between a head and a maximal projection, it is not obviously a *binding* relation (see Safir 2008 for discussion), though there are theories consistent with this view for local anaphora, and without going into details, Kratzer's theory is one of them. She proposes that all binding relations are effected by heads that induce property formation by lambda extraction, such that the specifier of the head must bind a variable in the sister to the head. For every binding relation, there is a head of this kind.

However, Agree was originally formulated as a relation between a head and a phrase, with morphological consequences in the case of subject-verb agreement. In many Icelandic cases, an oblique argument is bound by a non-nominative subject, while the verb in the same clause agrees with a Nominative non-subject, as illustrated in (A3a–b).

(A3) a. Henni þykir broðr sinn/*hennar leiðinlegar.
          she.DAT thinks brother SIN/her   boring
          'She finds her brother boring.'

      b. Konunginum voru gefnar ambáttir í höll sinni/?hans.
          the-king.DAT were given slaves   in palace SIN/his
          'The king was given slaves in his palace.'

                                                          (Zaenen *et al.* 1990: 102, 112)

In these examples, the structural subject is Dative, so the verb agrees with the post-verbal Nominative. The possessive anaphor *SIN* in Icelandic (inflected for agreement with what it modifies), which must normally be bound by a structural subject, is bound by the Dative subject. If Agree is crucially identifying which elements may move, then it is already failing to select between the possibilities in this case, since it is the Dative that fills the structural subject position (presumably to satisfy EPP). Moreover, it is also clear that verb agreement is not establishing the anaphoric relation (since the verb usually agrees with a structural subject that is Nominative). One could add an additional ad hoc head H$_{ANAPHOR}$ that forms an agreement relation between the Dative subject and something in the complement of that head, but this would be an antecedent agreement relation, not verb agreement.

In short, if Agree on a functional node determines verb agreement, then a different functional node must be responsible for anaphor-binding, with the result that all that is left of Agree is that it is a kind of local c-command mapped onto one interpretive or morphological relation or another. Rather than giving Agree rhetorical priority, it seems more natural to see agreement relations as one

of the family of local interpretive relations mapped onto c-command and bounded by intervention, which is exactly the theory proposed here. It is too great a task to provide an alternative account of local anaphora here, but one such alternative theory might be based on prominence relations within a domain that is defined by intervention, a model similar to that suggested for Case. In such an account, the role of heads may very well emerge as compatible with an account like Kratzer's, but the key notion is then c-command for an interpretive relation—not operational Agree.

A more interesting argument for Agree as the effector of anaphoric relations is presented by Reuland (2005) and adopted by Chomsky (2007a, 2008). Reuland argues that Agree is crucial to anaphoric relations because there are cases where a probe can effect an anaphoric interpretation when the antecedent of an anaphor is not c-commanded by its antecedent. In these cases, it is argued, the probe c-commands both anaphor and antecedent locally. The argument clearly does not go through for English.

(A4) a. There appear to be **a lot of gifted children** getting advantages that others deserve.

  b. ?*A lot of gifted children* seem to *each other's* parents to be getting advantages that others deserve.

  c. *There appear to *each other's* parents to be **a lot of gifted children** getting advantages that others deserve.

In these instances, an Agree relation is supposed to hold between T and the bolded nominal, but where *each other's* and *a lot of gifted children* are co-construed, the presence of T agreeing with both nominals in (A4c) is not enough to effect anaphora. However, the arguments that Reuland makes are for Norwegian and Icelandic.

Reuland suggests that (A5a) in Norwegian provides the right configuration and I have added (A5b), which does not require an awkward context, but shows the same effect.

(A5) a. Det ble   introdusert *en mann* for   *seg selv/*ham selv.*   Norwegian
          it   became introduced   a man   to   SEG-self
          'A man was introduced to himself.'

  b. Det ble   introdusert *en mann* for   læreren   *sin/*hans.*
          it   became introduced   a man   to   teacher-the SIN/his
          'A man was introduced to his teacher'

In these examples, the form of the anaphors *seg selv* and *sin* appear to indicate that they are bound by a structural subject, yet the only candidate antecedent is expletive *det*, which, Reuland assumes (and I concur) could not bind the subject-oriented antecedent even if *en mann* does c-command *seg selv*. For data such as these, however, it appears that the choice of verb and idiolect are crucial. The consultant whose judgments Reuland reports, Øystein Nilsen, also accepts (A6) (thanks to Øystein Nilsen, p.c., for supplying the Norwegian examples in this

section and his enlightening discussion of them).

(A6) a.   Psykiateren   introduserte en forvirret mann for   seg/*ham selv.
             psychiatrist-the introduced   a confused man to   SEG/him self
             'The psychiatrist introduced the confused man to himself.'
       b.   Rektoren   introduserte   student-en/en student for
             headmaster-the introduced   student-the/a   student to
             læreren   [sin/hans].
             teacher-the   SIN/his
             'The headmaster introduced a/the student to his teacher.'

Øystein Nilsen (p.c.) remarks:

> The grammatical version of [(A6)] with *seg* is ambiguous for me, even
> though the subject-bound construal requires perhaps a confused shrink,
> rather than a confused man […]. Other Norwegians have a strictly [subject]-
> oriented *seg selv*, so I may be 'special' in not requiring subject-orientation.

He reports the same ambiguity for *sin* in (A6b), but *hans* cannot refer to either of
the potential antecedents. If the active form of (A5a) in (A6) permits the direct
object to bind the *seg self* or *sin* anaphors, then there is no reason to suppose that
Agree is crucial to permitting the 'subject oriented' form in this case. Unless an
additional Agree relation is added here for the complementation of 'introduce' in
Nilsen's dialect (and incidentally, Nilsen's dialect does not even show subject-
verb agreement in (A5)–(A6)), the agreement theory seems to be exactly wrong,
and if such a relation is added, then no evidence for the Agree theory of anaphor
antecedence is to be gained by it.[46]

---

[46]   Nilsen suggests (and in this note I paraphrase what he told me) that *introdusere* is special, in
        that other ditransitives allow both the pronominal and the anaphoric possessor in examples
        parallel to (A6b).

        (i)   Jeg sendte Jens til faren   sin/hans.   *Norwegian*
               I   sent   Jens to father-the SIN/his
               'I sent Jens to his father.'

        He notes that these other ditransitives allow for VP topicalization, leaving the dative PP
        behind. In that case, only the pronominal possessor is possible. *Introdusere* doesn't allow
        such partial VP fronting in the first place (the verb in VP topicalizations in Norwegian
        retains finite inflection).

        (ii)   [Sendte Jens] gjorde jeg til faren   hans/*sin.
               sent   Jens   did   I   to father-the his/SIN

        (iii) *Introduserte Jens   gjorde jeg for Per.
               introduced   Jens   did   I   for Per

        Nilsen interprets this to mean that the PP in *send*-type VPs is allowed in two different
        hierarchical positions, one 'adjoined' outside the constituent made up by the verb and the
        direct object, and another, 'Larsonian' position, c-commanded by the direct object.
        Introdusere would only allow the PP to occur in the lower, Larsonian position. Only the
        higher position would be able to feed partial VP fronting of the relevant sort, and possessive
        SIN bound by the object would only occur in the Larsonian position, while possessive 'his'

The purported Icelandic evidence Reuland presents is based on (86) elicited from Halldor Sigurðsson (p.c.).

(A7) Thað   kom   maður   með börnin sin/*hans.   *Icelandic*
      there   arrived   a-man   with   children SIN/his
      'A man arrived with his children.'

Reuland suggests that this is another example where the *sin* anaphor is only possible because of antecedency facilitated by Agree, not by the expletive *thað* or c-command from *maður*. However, the use of *sin* is freer in Icelandic than it is in Norwegian, and it is not always the case that it requires a structural subject, as in (A8), pointed out to me by Halldor Sigurðsson (p.c.).

(A8) Eftir vinnu var bara farið heim til sín.
      after   work   was just   gone   home to   self
      'After work you/they/we/people just went home (to X-selves).'

For (A8) one might argue that Agree is doing the work in the absence of an antecedent, or perhaps that Agree has no role to play at all since a structural antecedent is not always required. In any case, examples like (A7) are not strong evidence for Agree as the effector of anaphoric relations, and the English and Norwegian examples seem to suggest that is not.

These arguments that Agree can replace Principle A are not persuasive, but even if there are other such arguments, the need for c-command to regulate resumption, scope intervention effects, and logophoric pronoun distributions remains unchallenged. The case for positing a structural relation more general than Agree as part of HLF seems secure.

Appendix B: The Varieties of Movement

VSA does not permit all the varieties of movement that Pesetsky (2000) has argued must exist. In particular, VSA does not permit any sort of covert phrasal movement, insofar as the highest occurrence of a non-terminal must be pronounced high. The question that must be addressed is whether or not it is necessary to weaken (51) or to seek different sorts of accounts for contexts where covert phrasal movement has been argued to exist.

Pesetsky (2000) argues that covert phrasal movement is necessary to account for antecedent-contained deletion within the family of analyses stemming from May (1985). These analyses appeal to covert phrasal movement of a quantified relative clause (adjunction to VP) to avoid infinite regress in the italicized portion in (B1a) that is not pronounced.

---

would be restricted to the higher adjunct position. Nilsen's reasoning here seems to justify the basic c-command story and would require ancillary stipulations in an Agree-based approach.

(B1) a. George trusts everyone (who) you do *trust everyone...*
      b. T … [$_{VP}$ [$_{DP}$ everyone (who) you do [__]] [$_{VP}$ trust [*e*]]]

The VP [*trust* e] is then used to fill in the ellipsis. However, given copy theory, the 'empty' object of *trust* is still present as an infinitely regressing copy of what has moved (*everyone who you do*), and so Fox (2002) develops an alternative account whereby the clause portion of the relative is 'late attached' (consistent only with Revised Extension as proposed here). Without going into details, Fox's analysis of this construction does not appear to require any covert phrasal movement to take place because the relative clause coda with the ellipsis, attached late, is pronounced in its first-merged position as in (B2a), and is (discontinuously) associated with the relative clause nucleus (*everyone*).

(B2) a. George [*every* [[trusts [*every*one]] [who you do__]]]
      b. [*everyone* … [*who* [C+WH-REL [*George* [T [*who* [*George* [$_A$ *v** [V *who*]]]

Interpreting Fox's approach slightly, the portion after *(who) you do*, which is elided at the bracket marked 'A', can be 'filled in' with the interpreted contents of the lower VP (*trust everyone*, where *everyone* must be the interpreted content of *who*). See Fox (2002) for discussion.

Pesetsky also posits a correlation between covert phrasal movement and the presence of superiority effects, such that *wh-in-situ* phrases undergo covert phrasal movement where superiority effects are present. Superiority is neutralized when the *wh*-phrases are D(iscourse)-linked.

(B3) a. Who saw who?
      b. * Who did who see?
      c. Which person did which person see?

It is not clear to me whether or not head movement pronounced low in VSA is empirically equivalent to Pesetsky's 'feature movement', but the optimal analytic result from the VSA perspective would be to argue that head/feature movement is the same for both D-linked and non-D-linked *wh-in-situ* and the superiority effect is a result of an intervention or interpretive condition neutralized by D-linking (structurally or semantically) (such as a choice function; e.g., Dayal 2002 or Reinhart 2006).

With respect to the linguistic variation in the realization of multiple interrogation structures, Pesetsky posits different complementizers that are stipulated to trigger single or multiple overt movement (or neither). These stipulations are only attractive if they are embedded in the rich set of assumptions that Pesetsky employs in order to draw together generalizations about movement, and so it is not clear that any real translation of his theory into VSA is possible. Richards (1999) and Pesetsky appeal to 'Shortest Move' which favors iMerge of a closer *wh*-Q-phrase rather than a more distant one to satisfy a trigger, but since I do not permit operational triggers, I cannot appeal to such a distinction. Moreover, Shortest Move only works in Pesetsky's system if covert phrasal movement unrelated to the ACD analysis exists, which is not established independent of

superiority effects. In VSA, the superiority effect can only arise as an intervention effect, such as the one described in (B4).

(B4) *Superiority as Intervention*[47]

The *wh*-phrase Q_{WH1} that is assigned WH_{D1} cannot have a lowest occurrence that is intervened by an occurrence of a Q_{WH2} if Q_{WH2} shares the scope of the Q_{WH1}.

For Q_{WH2} to share the scope of Q_{WH1}, it is necessary to introduce a way in which two quantifiers could share the same domain assignment. In VSA, scope assignment occurs at the unique point in a derivation where the Q or QP first c-commands WH_D, where WH_D is the domain set by the C_{WH} complementizer. Since domain assignment is unique, I need to add a device that allows for two quantifiers to share a domain for (83) to make sense. Notice that VSA permits 'tucking in' as in Richards (1999).

(B5)
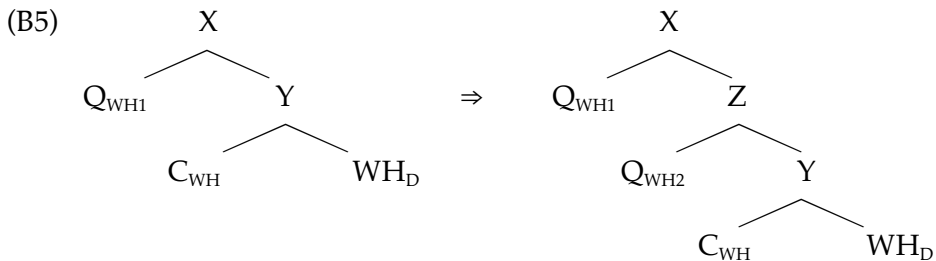
In such configurations, an interpretive rule must be introduced to permit the scope of Q_{WH2} to be interpreted as parasitic on the scope of Q_{WH1}. A solution based on an instance of subMerge in (B5) has the virtue of preserving the primacy of initial wh-movement, while (B4) captures the relevant phenomenon in terms of intervention. If so, tucking in is then movement to a (parasitic) criterial position. A full treatment of superiority, one that reduces it to intervention without appealing to covert phrasal movement, is beyond the scope of this article, but these remarks suggest a plausible avenue to explore within VSA.[48]

---

[47] Hornstein (2001) has suggested that superiority may be a form of weak crossover, and a correlation between the two phenomena has been established, for example, by Adesola (2006). If so, a dependency relation holds between the extracted wh-phrase and the in situ one, and the intervention effect may be a subcase of the Independence Principle, proposed in Safir (2004b), where it is argued that a variable or the constituent that contains it cannot c-command anything the variable depends on.

[48] Challenges for VSA remain. Where phrasal movement is required, as in English, I have already accepted the (inelegant) possibility that a head can require that it be assigned to a non-terminal node (65). Nothing in the system requires multiple fronting, and indeed that is not stateable in terms of how a head is subMerged. Languages such as Bulgarian that require multiple fronting in multiple interrogations will have to be addressed in some other way. It could be that whatever permits multiple phrases to be assigned the same scopal domain can also be manipulated to express these differences (though many typological issues, including those raised by Bošković 2002, need consideration). Such cases pose a challenge to VSA, but they are problems that seem tractable insofar as appeal to triggers distinguishing iMerge from eMerge are not required.

References

Ackema, Peter & Ad Neeleman. 2004. *Beyond Morphology: Interface Conditions on Word Formation*. Oxford: Oxford University Press.

Adesola, Oluseye. 2006. On the absence of superiority and weak crossover effects in Yoruba. *Linguistic Inquiry* 37, 309–318.

Baker, Mark & Nadya Vinokurova. 2008. Two modalities of case assignment in Sakha. Ms., Rutgers University.

Beck, Sigrid & Shin-Sook Kim. 1997. On *wh-* and operator scope in Korean. *Journal of East Asian Linguistics* 6, 339–384.

Beck, Sigrid. 1996. Quantified structures as barriers for LF-movement. *Natural Language Semantics* 4, 1–56.

Beghelli, Filippo & Tim Stowell. 1997. Distributivity and negation. In Anna Szabolcsi (ed.), *Ways of Scope Taking*, 71–107. Dordrecht: Kluwer.

Bhatt, Rajesh & Roumyana Pancheva. 2004. Late merger of degree clauses. *Linguistic Inquiry* 35, 1–45.

Bittner, Maria & Ken Hale. 1996. The structural determination of Case and agreement. *Linguistic Inquiry* 27, 1–68.

Bobaljik, Jonathan & Samuel Brown. 1997. Interarboreal operations: Head movement and the extension requirement. *Linguistic Inquiry* 28, 345–356.

Bobaljik, Jonathan & Susi Wurmbrand. 2005. The domain of agreement. *Natural Language and Linguistic Theory* 23, 809–865.

Bobaljik, Jonathan & Susi Wurmbrand. 2008. Case in GB/minimalism. In Andrej Malchukov & Andrew Spencer (eds), *The Oxford Handbook of Case*, 44–58. Oxford: Oxford University Press.

Bobaljik, Jonathan. 2002. A-chains at the PF-interface: Copies and 'covert' movement. *Natural Language and Linguistic Theory* 20, 197–267.

Bobaljik, Jonathan. 2008. Where's phi? Agreement as a post-syntactic operation. In David Adger, Susana Béjar, & Daniel Harbour (eds.), *Phi-Theory: Phi-Features across Interfaces and Modules*, 295–328. Oxford: Oxford University Press.

Boeckx, Cedric & Norbert Hornstein. 2006. The virtues of control as movement. *Syntax* 9, 118–130.

Boeckx, Cedric. 2003. *Islands and Chains*. Amsterdam: John Benjamins.

Borer, Hagit. 2005a. *In Name Only: Structuring Sense* vol. I. Oxford: Oxford University Press.

Further issues arise if the *in-situ* strategy for multiple wh-interrogatives is not uniform across languages. For example, Dayal (2002) contends that *in situ wh* in islands in Hindi only permits single pair answers to questions, not multiple pair answers. This difference does not follow from assuming that the Hindi *in situ* strategy is head movement pronounced low (Dayal proposes that islands require interpretation by choice function). Dayal attributes this locality effect from a (stipulated) clausemate restriction on covert *wh*-movement. The effect might be recast as an intervention effect if multiple interrogation domains are linked to event or tense domains in Hindi, such that the head setting one event/tense domain intervenes for another (a similar domain limits anaphora in Hindi, see Safir, 2004a: 166–170). Then multiple pair interpretations would be limited to a domain containing a single event/tense, with single pair interpretations possible where an event/tense head intervenes between criterial landing sites. Head movement would not then have to be restricted in an ad hoc way. The exploration of this possibility is beyond the scope of this essay.

Borer, Hagit. 2005b. *The Normal Course of Events: Structuring Sense*, vol. II. Oxford: Oxford University Press.

Bošković, Željko. 2002. On multiple *wh*-fronting. *Linguistic Inquiry* 33, 351–383.

Bošković, Željko. 2007. On the locality and motivation of Move and Agree: An even more minimal theory. *Linguistic Inquiry* 38, 589–644.

Bossong, Georg. 1985. *Differentielle Objektmarkierung in den neuiranischen Sprachen*. Tübingen: Gunter Narr Verlag.

Cable, Seth. 2008. Towards the elimination of pied-piping: Evidence from Tlingit. Ms., University of Massachusetts, Amherst.

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

Chomsky, Noam. 1986. *Barriers*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1995. *The Minimalist Program.* Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 1–52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3, 104–131. New York: Oxford University Press.

Chomsky, Noam. 2007a. Approaching UG from below. In Uli Sauerland & Hans-Martin Gärtner (eds.), *Interfaces + Recursion = Language? Chomsky's Minimalism and the View from Syntax–Semantics* (Studies in Generative Grammar 89), 1-29. Berlin: Mouton de Gruyter.

Chomsky, Noam. 2007b. Of minds and language. *Biolinguistics* 1, 9–27.

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

Christiansen, Morten & Nick Chater. 2008. Language as shaped by the brain. *Behavioral and Brain Sciences* 31, 489–558.

Clements, George. 1975. The logophoric pronoun in Ewe: Its role in discourse. *Journal of West African Languages* 10, 141–177.

Collins, Chris. 2002. Eliminating labels. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program*, 43–64. Malden, MA: Blackwell.

Dayal, Veneeta. 2002. Single pair vs. multiple pair answers, *wh*-in situ and scope. *Linguistic Inquiry* 33, 512–520.

Epstein, Samuel David, Erich M. Groat, Ruriko Kawashima & Hisatsugu Kitahara. 1998. *A Derivational Approach to Syntactic Relations.* New York: Oxford University Press.

Fox, Danny & David Pesetsky. 2005. Cyclic linearization of syntactic structure. *Theoretical Linguistics* 31, 1–45.

Fox, Danny. 2002. ACD and the copy theory of movement. *Linguistic Inquiry* 33, 64–96.

Fox, Danny. 2003. On logical form. In Randall Hendrick (ed.), *Minimalist Syntax*, 82–123. Malden, MA: Blackwell.

Frank, Robert & K. Vijay-Shanker. 2001. Primitive c-command. *Syntax* 4, 164–204.

Haegeman, Liliane & Raffaella Zanuttini. 1996. Negative concord in West Flemish. In Adriana Belletti & Luigi Rizzi (eds.), *Parameters and Functional Heads: Essays in Comparative Syntax*, 117–179. New York: Oxford University Press.

Haegeman, Liliane. 1995. *The Syntax of Negation*. Cambridge: Cambridge University Press.

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

Heck, Fabian. 2004. A theory of pied piping. Tübingen: Eberhard Karls Universität Tübingen dissertation.

Heck, Fabian. 2007. On certain properties of pied piping. Ms., Universität Leipzig.

Hornstein, Norbert & Jairo Nunes. 2008. Adjunction, labeling, and bare phrase structure. *Biolinguistics* 2, 57–86.

Hornstein, Norbert. 1999. Movement and control. *Linguistic Inquiry* 30, 69–96.

Hornstein, Norbert. 2001. *Move! A Minimalist Theory of Construal*. Malden, MA: Blackwell.

Hornstein, Norbert. 2003. On control. In Randall Hendrick (ed.), *Minimalist Syntax*, 6–81. Malden, MA: Blackwell.

Huang, James. 1992. Logical relations in Chinese and the theory of grammar. Cambridge, MA: MIT dissertation.

Jayaseelan, K.A. 2008. Bare phrase structure and specifierless syntax. *Biolinguistics* 3, 87–106.

Jónsson, Jóhannes. 1996. Clausal architecture and case in Icelandic. Amherst, MA: University of Massachusetts dissertation.

Kim, Hyunjoo. Forthcoming. Intervention effects and movement in *wh*-in-situ constructions. New Brunswick, NJ: Rutgers University dissertation.

Koster, Jan. 1976. *Locality Principles in Syntax*. Dordrecht: Reidel.

Kratzer, Angelika & Junko Shimoyama. 2002. Indeterminate pronouns: The view from Japanese. *Proceedings of the 3rd Tokyo Conference on Psycholinguistic*, 1–25. [Available online at the Semantics Archive: http://www.semantics archive.net/Archive/WEwNjc4Z.]

Kratzer, Angelika. 2009. Making a pronoun: Fake indexicals as windows into the properties of pronouns. *Linguistic Inquiry* 40, 187–237.

Landau, Idan. 2000. *Elements of Control: Structure and Meaning in Infinitival Constructions*. Dordrecht: Kluwer.

Landau, Idan. 2003. Movement out of control. *Linguistic Inquiry* 34, 471–498

Landau, Idan. 2006. Severing the distribution of PRO from case. *Syntax* 9, 155–170.

Lasnik, Howard & Juan Uriagereka with Cedric Boeckx. 2005. *A Course in Minimalist Syntax*. Malden, MA: Blackwell.

Lebeaux, David. 1991. Relative clauses, licensing and the nature of the derivation. In Susan Rothstein (ed.), *Perspectives on Phrase Structure Heads and Licensing*, 209–239. New York: Academic Press.

Legate, Julie Anne. 2008. Morphological and Abstract Case. *Linguistic Inquiry* 39,

55–102.

Linares, Carlo. 2008. Approaching DOM in Viable Syntax. Manuscript, Rutgers.

Linares, Carlo. Forthcoming. On the typology of movement [tentative title]. New Brunswick, NJ: Rutgers University dissertation.

Longobardi, Giuseppe. 1994. Reference and proper names: A theory of N-movement in syntax and logical form. *Linguistic Inquiry* 25, 609–665.

Marantz, Alec. 1991. Case and licensing. *ESCOL '91*, 234–253. [Reprinted 2000 in Eric Reuland (ed.), *Arguments and Case: Explaining Burzio's Generalization*, 11–30. Amsterdam: John Benjamins.]

Matushansky, Ora. 2006. Head movement in linguistic theory. *Linguistic Inquiry* 37, 69–109.

May, Robert. 1985. *Logical Form*. Cambridge, MA: MIT Press.

McCloskey, James. 1979. *Transformational Syntax and Model-Theoretic Semantics: A Case Study in Modern Irish*. Dordrecht: Reidel.

McCloskey, James. 2002. Resumption, successive cyclicity and the locality of operations. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and the Explanation in the Minimalist Program*, 184–226. Malden, MA: Blackwell.

McFadden, Thomas. 2004. The position of morphological case in the derivation: A study on the syntax-morphology interface. Philadelphia, PA: University of Pennsylvania dissertation.

Neeleman, Ad, & Hank van de Koot. 2002. The configurational matrix. *Linguistic Inquiry* 33, 529–574.

Nunes, Jairo. 1999. Linearization of chains and phonetic realization of chain links. In Samuel David Epstein & Norbert Hornstein (eds.), *Working Minimalism*, 217–249. Cambridge, MA: MIT Press.

Nunes. Jairo. 2004. *Linearization of Chains and Sideward Movement*. Cambridge, MA: MIT Press.

Obenauer, Hans. 1984. On the identification of empty categories. *The Linguistic Review* 4: 153–202.

Pesetsky, David & Esther Torrego. 2001. T-to-C movement: Causes and consequences. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language*, 355–426. Cambridge, MA: MIT Press.

Pesetsky, David. 2000. *Phrasal Movement and Its Kin.* Cambridge, MA: MIT Press.

Ramchand, Gillian. 2008. *Verb Meaning and the Lexicon: A First Phase Syntax*. Cambridge: Cambridge University Press.

Reinhart. Tanya. 2006. *Interface Strategies*. Cambridge, MA: MIT Press.

Reuland, Eric. 2005. Agreeing to bind. In Hans Broekhuis, Norbert Corver, Riny Huybregts, Ursula Kleinerz, & Jan Koster (eds.), *Organizing Grammar: Studies in Honor of Henk van Riemsdijk*, 505–513. Berlin: Mouton de Gruyter.

Richards, Norvin. 1999. Featural cyclicity and the ordering of multiple specifiers. In Samuel David Epstein & Norbert Hornstein (eds.), *Working Minimalism*, 127–158. Cambridge, MA: MIT Press.

Rizzi, Luigi. 1990. *Relativized Minimality*. Cambridge, MA: MIT Press.

Rizzi, Luigi. 2006. On the form of chains: Criterial positions and ECP effects. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *WH-Movement: Moving On*, 97–133. Cambridge, MA: MIT Press.

Ross, John R. 1967. Constraints on variables in syntax. Cambridge, MA: MIT

dissertation.

Safir, Ken. 1986. Relative clauses in a theory of binding and levels. *Linguistic Inquiry* 17, 663–689.

Safir, Ken. 1987. The syntactic projection of lexical thematic structure. *Natural Language and Linguistic Theory* 5, 561–601.

Safir, Ken. 1999. Vehicle change and reconstruction in A'-chains. *Linguistic Inquiry* 30, 587–620.

Safir, Ken. 2004a. *The Syntax of Anaphora*. Oxford: Oxford University Press.

Safir, Ken. 2004b. *The Syntax of (In)dependence*. Cambridge, MA: MIT Press.

Safir, Ken. 2005. Introducing exxtension. In Yukio Otsu (Ed.), *The Proceedings of the Fifth Tokyo Conference on Psycholinguistics*, 1–31. Tokyo: Hitsuzi Syobo.

Safir, Ken. 2008. Coconstrual and narrow syntax. *Syntax*: 330–355.

Safir, Ken. In preparation. Irish complementizers and extraction without triggers. Ms., Rutgers University.

Sauerland, Ulrich & Fabian Heck. 2003. LF-intervention effects in pied-piping. *NELS* 33, 347–366.

Seely, T. Dan. 2006. Merge, derivational c-command, and subcategorization in a label free syntax. In Cedric Boeckx (ed.), *Minimalist Essays*, 182–217. Amsterdam: John Benjamins.

Starke, Michal. 2004. On the inexistence of specifiers and the nature of heads. In Adriana Belletti (ed.), *Structures and Beyond: The Cartography of Syntactic Structures*, vol. 3, 252–268. New York: Oxford University Press.

Stepanov, Arthur. 2001. Late adjunction and minimalist phrase structure. *Syntax* 4, 94–125.

Torrego, Esther. 1998. *The Dependencies of Objects*. Cambridge, MA: MIT Press.

van Urk, Coppe. 2009. On obligatory control. Ms., University of Utrecht.

Vinokurova, N.1998. The internal structure of noun phrases in Sakha. Tromsoe: Universitetet i Tromsø M.Phil. thesis.

Watanabe, Akira. 2006. The pied piper feature. In Lisa Lai-Shen Cheng & Norbert Corver (eds.), *WH-Movement: Moving On*, 47–70. Cambridge, MA: MIT Press.

Zaenen, Annie, Joan Maling & Höskuldur Thráinsson. 1990. Case and grammatical relations: The Icelandic passive. In Joan Maling & Annie Zaenen (eds.), *Modern Icelandic Syntax*, 95–136. New York: Academic Press. [Reprinted from *Natural Language and Linguistic Theory* 3, 441–483, 1985].

Ken Safir
Rutgers University
Department of Linguistics
18 Seminary Place
New Brunswick, NJ 08901
USA
safir@ruccs.rutgers.edu

On Perceived Conceptual and Methodological Divergences in Linguistic Theory and Cognitive Science: Distributional Analyses, Universal Grammar, and Language Acquisition

Garrett Neske

Many of the theoretical innovations in linguistic theory and cognitive science due to Noam Chomsky, and now considered to be cornerstones of the biolinguistic/*I*-language approach, are often viewed as radical replacements of earlier views in structural linguistics and psychology. In particular, the story often goes that Chomsky abandoned the structuralist tradition, along with "discovery procedures" and distributional analyses as a means toward understanding linguistic structure, in his development of transformational generative grammar. The position is also assigned to Chomsky that the study of language should consist of the elucidation of the innate component of the human mind and brain that allows for the acquisition of the rule systems of natural language and the real-time parsing of linguistic data. Both of these supposed truisms, however, require a serious reevaluation if the foundations of biolinguistics are to guide future research in the proper direction. Any divergences of generative grammar in linguistic theory and of universal grammar (UG) in cognitive science from previous conceptions of language and the mind have been too exaggerated. A careful consideration of the history and subsequent development of generative grammar and the biolinguistic/*I*-language approach will show, first, that distributional analyses were never abandoned in Chomsky's program and, second, that external linguistic data are integral to a theory of UG. These clarifications are absolutely essential if one is to make progress in biolinguistics.

Transformational generative grammar, as outlined in Chomsky (1957), detailed in Chomsky (1955, 1975a), and modified and refined several times throughout the second half of the last century (Chomsky 1965, 1972, 1975b, 1981, 1995) constituted a major departure from the structuralist tradition in linguistics. The aim of linguistic analysis from the structuralist, or distributionalist perspective, as exemplified in the work of Leonard Bloomfield (1933), Zellig Harris (1951), and others is to extract structure and constraints on linguistic representation from the distribution of formal objects such as phonemes. That is, the focus

should be on the likelihood of occurrence of a particular object given the occurrence of another object, with the aim of constructing hypotheses about the higher-level structure of language (Huck & Goldsmith 1986).

With the risk of oversimplifying, one may conceive of the distributional program as a bottom-up approach, akin to exploratory data analysis and clustering, in which the inference of structure comes from the distributional relations of primitive elements in a corpus, while Chomsky's transformational generative grammar, in contrast, takes a top-down approach, in which hypotheses about structure come from systems that generate a corpus. Working through many sentential examples, Chomsky became unconvinced that a distributional program could accomplish the task of inferring higher-level structure from the relations between elements. Chomsky's solution was a transformational generative grammar that could in principle generate the discrete infinity of grammatical expressions in a natural language. Furthermore, the 'transformational' component of this generative grammar differed markedly from the distributional program in positing different levels of representation connected by transformational rules. One level, deep structure in Chomsky's earliest Standard Theory (Chomsky 1957) and logical form in his most recent Minimalist Program (Chomsky 1995), maps syntactic structure to semantic interpretation and transformation rules use this level as input to yield an output of another level of representation, surface structure (Chomsky 1957) or phonetic form (Chomsky 1995).

In considering the birth of generative grammar and its subsequent development, one must be careful to consider the elements of earlier linguistic analysis that have been retained. Chomsky's generative approach was never completely divorced from distributionalism. Chomsky realized that purely distributional procedures could not, as an empirical matter, yield higher-level language structures from a corpus, which was the initial impetus for his construction of a top-down, meta-theory of language (Chomsky 1951). He did not, however, reject the distributional program altogether as a necessary component of the theory of grammar. In the years prior to the publication of Chomsky (1957), the new field of information theory introduced a principled, mathematical description of coding and transmission over a noisy channel (Shannon 1948), which, for Chomsky (1955, 1975a) and later for Harris (1991), seemed to bear on linguistic analysis. Shannon (1951) showed that information theory provided a way to predict letters in a text given the previous letters. Considering information-theoretic analysis as an empirical approach to the study of language structure, Chomsky notably worked with information theory pioneer Peter Elias to develop clustering algorithms for syntactic category formation and included the results of this analysis and a discussion of the scope and limits of distributional procedures in Chomsky (1955, 1975a: chap. V in particular). While Chomsky (1955, 1975a) is primarily an argument for a generative theory of language, there are several instances in which it is clear that a wholesale rejection of distributional, bottom-up methods in linguistic analysis is not the correct position to take, and in fact that a distributional approach might be indispensable for syntactic category formation:

> Note that there is no question being raised here as to the legitimacy of the probabilistic approach, just as the legitimacy of the study of meaning was in no way brought into question when we pointed out […] that projection can-

not be defined in semantic terms. Whether or not the statistical study of language can contribute to grammar, it surely can be justified on quite independent grounds. These three approaches to language (grammatical, semantic, statistical) are independently important. In particular, none of them requires for its justification that it lead to solutions for problems which arise from pursuing one of the other approaches.   (Chomsky 1975a: 148, fn. 19)

Thus, one would be too swift to regard Chomsky's approach as a successor to an extinct distributional program. Whether by lack of careful consideration of the foundations of generative theory or by trying to fit generative theory into some kind of Kuhnian paradigm shift, it is too often considered a truism that discovery procedures implemented to build categories and structures from primitive elements in linguistic data differ irrevocably from Chomsky's linguistics (Searle 1972). For one, the distributionalist program has been used to great effect by computational and corpus linguists for parsing, machine translation, and other tasks 'outside' of the biolinguistic/*I*-language approach. Harris's early hypothesis that morpheme boundaries corresponded to measurable variations in the complexity of phoneme sequences (Harris 1955) can be tested on a corpus of utterances (e.g., Hayes & Clark 1970, Tanaka-Ishii & Jin 2006), with the prediction that local phonemic entropy maxima will occur at morpheme boundaries. Yet, the most important point is that the distributional approach is not irrelevant to *I*-language, in which it is often assumed that the only relevant theory for language structure is generative and syntacto-centric (Jackendoff 1998). While distributionalist methods might fail at inducing syntactic structure, they are useful, and no doubt essential, to the segmentation problem (i.e. the determination of syntactic categories from linguistic data).

As a case in point, the famous example sentence in Chomsky (1955, 1957, 1975a), *Colorless green ideas sleep furiously*, is often touted as emblematic of the failure of distributional methods to induce structure from linguistic data. To a certain extent, this is true; the probability that each word in this sentence follows the other is thought to be vanishingly small, such that a parser would classify such a sentence as ungrammatical. Nevertheless, while the probability of the sentence occurring as written might be the same as that of the 'word salad' one obtains in reading it backwards, the sentence is pronounced with normal intonation and judged as grammatical by virtue of being in the class of sentences of the form Adjective–Adjective–Noun–Verb–Adverb (Chomsky 1975a: 145–147). A generative theory of language would require the segmentation of linguistic data into higher-level categories that serve as elements of the syntactic structure. That certain, modern distributional methods lead to a high probability of occurrence of *Colorless green ideas sleep furiously* compared to the reverse string (e.g., Pereira 2000) is actually consistent with the research program discussed in Chomsky (1955, 1975a); while distributional methods are not sufficient to induce syntactic structure, they are indispensable for the delineation of syntactic categories in linguistic data.

The birth of transformational generative grammar was intimately tied to the so-called cognitive revolution that rejected behaviorism (Chomsky 1959) and emphasized the role of innate cognitive architecture in producing behaviors. The position of *nativism* is often associated with the logical problem of language

acquisition and the argument from the poverty of the stimulus: How does the child learn the grammar of its native language from degenerate and limited linguistic data without a pre-existing structure specialized for the task? The guiding principle for understanding this structure is the notion of UG, a theory of the biologically instantiated ability to acquire and utilize the rule systems of a natural language (Chomsky 1965, 2006).

Given that linguistic structures exhibit a complexity not observed in any non-human communication system, or in any other behavior for that matter, it is almost obligatory to infer that there is an aspect of human biology and cognition, not shared with any non-human, responsible for the potential for acquisition of these structures and their use in cognition and communication. UG is a system that describes this unique ability. What kind of 'system' is UG posited to be? In the biolinguistic/*I*-language approach, UG is not a set of linguistic principles external to the individual like the grammars of specific natural languages, but an internal system for acquiring any natural language. In another sense, UG can be considered the initial state of the language faculty before any language-specific input. Thus, the focus of the biolinguistic/*I*-language approach is on 'human language' as opposed to the study of specific natural languages and associated corpora (i.e. *E*-language). This point is central to the later Principles–and–Parameters (P&P) approach to language acquisition (Chomsky & Lasnik 1993). The P&P approach holds that the individual has instantiated knowledge (though of course not 'conscious' knowledge) of fundamental principles of linguistic operation that are necessary components of all natural languages and that natural languages manifest their differences by setting parameters in the existing UG (i.e. the initial state of the language faculty).

Unfortunately, as in the perceived distributional/generative divergence in the analysis of linguistic structures, a similar artificial rift is drawn between UG and statistical learning, the build-up of language structure from the statistical properties of a sound signal (Seidenberg *et al.* 2002). In fact, these perceived divergences are both undoubtedly derived from the urge to fit differing conceptions of language analysis and language acquisition into an internalist–externalist, or even nature–nurture, debate. As before, a careful consideration of the specific claims immanent in the biolinguistic/*I*-language approach will reveal the exaggeration that statistical learning is somehow an alternative account of language acquisition that is in conflict with UG.

To reiterate, UG is a theory of the initial state of the language faculty, which, in the P&P model, undergoes a setting of parameters driven by external linguistic data. This is a selectionist account of learning that, while a relatively recent viewpoint in cognitive science, has been prevalent in the study of the development of biological structures (Piattelli-Palmarini 1989). There is noticeable confusion in the literature, and thus often artificial criticisms, about the claims of UG and the biolinguistic/*I*-language approach. For one, UG does not entail the 'unlearnability' of language, at least not in the selectionist sense, as has unfortunately been the position assigned to it in the literature (e.g., Bates & Elman 1996, Seidenberg 1997). UG is surely based upon the argument from the poverty of the stimulus, appealing to an innate structure to compensate for the lack of evidence needed to acquire the rule systems of natural language, but this does not relegate

the stimulus to negligible status. While children are theorized to be born with a UG, linguistic stimuli must allow the UG to converge on the correct grammar.

Many arguments countering, or purporting to counter, Chomsky's notion of UG and the biolinguistic/*I*-language approach rely on the efficacy of statistical learning in segmentation of the sound signal in early cognitive development. The landmark study, most often cited in statistical theories of language acquisition, demonstrating the power of statistical learning in the segmentation problem is that of Saffran *et al.* (1996). In this study, the authors show that infants can learn the 'words' of an artificial grammar within only minutes of exposure to sound samples of this grammar. They hypothesize that word segmentation is computed from local minima in transitional probabilities (TP) between syllables, where $TP(A\rightarrow B) = P(AB)/P(A)$ with $P(AB)$ the probability of syllable B following A and $P(A)$ the total frequency (i.e. probability) of A in the corpus (Yang 2004). While the data from Saffran *et al.* (1996) and others establish the dexterity with which children and adults detect the statistical distribution of sounds, it is far from straightforward what this implies about the innate capacity to learn language, let alone that a system like UG may not be necessary, which some authors have suggested (Seidenberg 1997, Bates & Elman 1996). In fact, it appears more likely that statistical learning does not play a role in deriving structure, but rather provides the language learner with a way to establish appropriate segmentations of the auditory signal, and thus the correct representation of linguistic elements, yet without the assignment of structure (Peña *et al.* 2002, Endress *et al.* 2005). The relation of these arguments to those asserting the divergence between the generative and distributional approaches to language analysis is evident (Gleitman 2002, in fact, argues that distributional approaches, while having little to do with cognitive science traditionally, have nonetheless been coapted to the field of language acquisition), and in both cases the same mistake is made; in the generative/distributional case, it is maintained that transformational generative grammar has no use for bottom-up procedures working on external data, while in the UG/statistical learning case, it is maintained that UG does not require any demarcation of the relational properties in the sound signal.

Statistical learning would, at first glance, appear to refute the need for a system like UG and validate a general-purpose learning scheme. The fact is, however, that there are an infinite number of statistical items in a sound signal of which the learner can keep track. The learner must be able to attend to the significant statistical correlations, such as transitional probabilities, and not, for instance, the probability of one syllable rhyming with the next or the probability that two adjacent vowels are both nasal (Yang 2004). This requires that the learner be equipped with some structure prior to linguistic exposure that accounts for the bias toward certain aspects of auditory stimuli and the neglect of other aspects. In effect, most arguments against the existence of a system like UG implicitly assume some initial structure that facilitates or constrains statistical learning. The appropriate question is not whether UG or a system like it exists, but what are its properties such that the learner can attend to certain features of the sound signal and acquire a complex rule system. General statistical learning schemes are not alone sufficient to account for the acquisition of a grammar. There is a need for an internal structure to attend to certain features of auditory stimuli.

Besides engaging in the segmentation problem, it might be that statistical learning can play a role in convergence to the correct grammar during language acquisition, though this has not yet been investigated experimentally like the segmentation problem (see Kuhl & Rivera-Gaxiola 2008 for a recent review of feasible experimental techniques in language acquisition). One issue in UG that must be addressed is that of parameter setting: What exactly does it mean for external linguistic data to 'set' a 'parameter'? The answer might involve a probabilistic component. In one learning scheme, UG represents the hypothesis space of grammars and parameter setting would involve the discarding of hypotheses that are inconsistent with external linguistic data. This notion is problematic, however, given the developmental data, which suggest that parameter setting is gradual and not punctuational (Bloom 1993) and that at any given time, the child has a representation of many possible grammars, not just one (Crain & Pietroski 2001). A scheme consistent with these data would be for each grammar in the hypothesis space to have a probability that is either increased or decreased depending on consistency or inconsistency with the linguistic input (Yang 2004). This kind of probabilistic hypothesis testing has been proposed as a neural basis for certain decision tasks (Gold & Shadlen 2002). While this situation still involves UG, statistical learning plays dual roles: segmentation of the sound signal and gradual convergence to the correct grammar. Again, to say that UG is incompatible with statistical learning is severely mistaken in the same way as the claim that transformational generative grammar is incompatible with the distributional approach and does not incorporate any of its methods.

It is important for the advancement of the biolinguistic/I-language approach not to overplay the supposed methodological and conceptual distance between distributionalism/statistical learning on the one hand and transformational generative grammar/UG on the other. All of these approaches are necessary for a full account of the structure of natural language and the development of grammar in the individual. Without distributional methods, for instance, an objective account of the notions of 'word' or 'lexicon', both essential to transformational generative grammar (e.g., the Merge operation acting upon *lexical* items in the Minimalist Program of Chomsky 1995), would be severely lacking. On the cognitive-science side, while UG is viewed as a biologically instantiated template to which the grammar of a natural language must conform, the theory of UG and the P&P approach have very little to say about the mechanisms of this 'conforming'. As in the necessity of the distributional approach for inducing syntactic categories, statistical learning appears sufficient (necessary?) to segment the sound signal into primitive linguistic elements based upon the statistics inherent in the stimulus. The propensity to frame new ideas as revolutionary attacks against past research programs is strong, but must often be mitigated through a careful consideration of the claims inherent in these new ideas and the extent to which they build upon previous ones. With this prescription in hand, it should be clear that the transformational generative grammar/distributionalism and UG/statistical learning chasms are not as wide as they are purported to be. In fact, all of these concepts and methods are utilized to some extent in the modern biolinguistic/I-language approach to natural language structure and the development and use of the biological language faculty.

References

Bates, Elizabeth & Jeffrey Elman. 1996. Learning rediscovered. *Science* 274, 1849–1850.

Bloom, Paul. 1993. Grammatical continuity in language development: The case of subjectless sentences. *Linguistic Inquiry* 24, 721–734.

Bloomfield, Leonard. 1933. *Language*. New York: Henry Holt.

Chomsky, Noam. 1951. Morphophonemics of Modern Hebrew. Philadelphia, PA: University of Pennsylvania MA thesis.

Chomsky, Noam. 1955. *The Logical Structure of Linguistic Theory*. Cambridge, MA: Microfilm, MIT Humanities Library.

Chomsky, Noam. 1957. *Syntactic Structures*. The Hague: Mouton.

Chomsky, Noam. 1959. A Review of B.F. Skinner's *Verbal Behavior*. *Language* 35, 26–58.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1972. *Studies on Semantics in Generative Grammar*. The Hague: Mouton.

Chomsky, Noam. 1975a. *The Logical Structure of Linguistic Theory*. New York: Plenum Press.

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

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

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2006. *Language and Mind*, 3rd edn. Cambridge: Cambridge University Press.

Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In Joachim Jacobs, Arnim von Stechow, Wolfgang Sternefeld & Theo Vennemann (eds.), *Syntax: An International Handbook of Contemporary Research*, vol. 1, 506–569. Berlin: Walter de Gruyter.

Crain, Stephen. & Paul Pietroski. 2001. Nature, nurture and universal grammar. *Linguistics and Philosophy* 24, 139–186.

Endress, Ansgar D., Brian J. Scholl, & Jacques Mehler. 2005. The role of salience in the extraction of algebraic rules. *Journal of Experimental Psychology: General* 134, 406–419.

Gleitman, Lila R. 2002. Verbs of a feather flock together II: the child's discovery of words and their meanings. In Bruce E. Nevin & Stephen M. Johnson (eds.), *The Legacy of Zellig Harris: Language and Information into the 21st Century*, vol. 1, 209–229. Philadelphia: John Benjamins.

Gold, Joshua I. & Michael N. Shadlen. 2002. Banburismus and the brain: Decoding the relationship between sensory stimuli, decisions, and reward. *Neuron* 36, 299–308.

Harris, Zellig S. 1951. *Methods in Structural Linguistics*. Chicago: University of Chicago Press.

Harris, Zellig S. 1955. From phoneme to morpheme. *Language* 31, 190–222.

Harris, Zellig S. 1991. *A Theory of Language and Information: A Mathematical Approach*. Oxford: Oxford University Press.

Hayes, John R. & Herbert H. Clark. 1970. Experiments on the segmentation of an artificial speech analogue. In John R. Hayes (ed.), *Cognition and the Development of Language*, 221–234. New York: John Wiley and Sons.

Huck, Geoffrey J. & John A. Goldsmith. 1986. Distributionalist and mediationalist themes in the development of linguistic theory. *Linguistic Inquiry* 17, 265–299.

Jackendoff, Ray. 1998. The architecture of the language faculty: A neominimalist perspective. In Peter Culicover & Louise McNally (eds.), *The Limits of Syntax*, 19–46. New York: Academic Press.

Kuhl, Patricia & Maritza Rivera-Gaxiola. 2008. Neural substrates of language acquisition. *Annual Review of Neuroscience* 31, 511–534.

Pereira, Fernando. 2000. Formal grammar and information theory: Together again? *Philosophical Transactions of the Royal Society A* 358, 1239–1253.

Peña, Marcela, Luca L. Bonatti, Marina Nespor & Jacques Mehler. 2002. Signal-driven computations in speech processing. *Science* 298, 604–607.

Piatelli-Palmarini, Massimo. 1989. Evolution, selection and cognition: From "learning" to parameter setting in biology and in the study of language. *Cognition* 31, 1–44.

Saffran, Jenny R., Richard N. Aslin & Elissa L. Newport. 1996. Statistical learning by 8-month-old infants. *Science* 274, 1926–1928.

Searle, John R. 1972. Chomsky's revolution in linguistics. *The New York Review of Books* 18 (June 29), 16–24.

Seidenberg, Mark S. 1997. Language acquisition and use: learning and applying probabilistic constraints. *Science* 275, 1599–1603.

Seidenberg, Mark S., Maryellen C. MacDonald & Jenny R. Saffran. 2002. Does grammar start where statistics stop? *Science* 298, 553–554.

Shannon, Claude E. 1948. A mathematical theory of communication. *Bell System Technical Journal* 27, 379–423, 623–656.

Shannon, Claude E. 1951. Prediction and entropy of printed English. *Bell System Technical Journal* 30, 50–64.

Tanaka-Ishii, Kumiko & Zhihui Jin. 2006. From phoneme to morpheme: Another verification using a corpus. *Lecture Notes in Computer Science*, 4285, 234–244.

Yang, Charles D. 2004. Universal grammar, statistics or both? *Trends in Cognitive Sciences* 8, 451–456.

Garrett Neske
New York University
Department of Linguistics
10 Washington Place
New York, NY 10003
USA
gtn206@nyu.edu

★ REVIEWS ★

Specters of Marx:
A Review of *Adam's Tongue* by Derek Bickerton

Sergio Balari & Guillermo Lorenzo

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

> *Die Sprache ist so alt wie das Bewußtsein—*
> *die Sprache ist das praktische, auch für andre Menschen existierende,*
> *also auch für mich selbst erst existierende wirkliche Bewußtsein,*
> *und die Sprache entsteht, wie das Bewußtsein, erst aus dem Bedürfnis,*
> *der Notdurft des Verkehrs mit andern Menschen.*
> Karl Marx & Friedrich Engels, *Die deutsche Ideologie* (1846)

1. Introduction

Jacques Derrida entitled one of his books *Spectres de Marx*. Pun intended. Derrida was playing with the plural 'specters' to suggest that many ghosts of Marx might be haunting modern Western society, while, at the same time, he was reminding us of the opening words of the *Communist Manifesto*: "*Ein Gespenst geht um in Europa*". Well, it seems that as far back as 1995 he was quite right after all, and that the specters of Marx materialize themselves in the most unexpected places: Derek Bickerton has just invoked one of them.

Derrida's point was that some of these materializations were good. Not just good, but necessary. Unfortunately, Bickerton's invocation in the context of language origins does not appear to be as necessary—nor as good, for that matter—as it is, according to Derrida, in other areas of contemporary Western thought.

Ein Gespenst geht um in Evolutionslinguistik—the specter of communication; but now in a Marxist disguise.

Not that Bickerton appears to be aware of what his proposal actually boils down to. Quite to the contrary, he seems to be in absolute oblivion of this. He seems in fact to be pretty convinced that he is doing hardcore biology; but he is not. Indeed, if we were to limit ourselves to provide a short version of Bickerton's model, we would already have an off-the-shelf paragraph to offer, namely, Marx and Engels's quote above. But Bickerton needed a whole book to say more or less the same, so we'd rather pay attention to the details.

This will be our task for the rest of the review.

2.   Constructing Niches in the Pleistocene Savanna

Bickerton has buried six feet down his original ideas about language evolution. *Language & Species* (Bickerton 1990) has ceased to be the basic reference for his position, this book has now been overthrown by *Adam's Tongue* (AT). So, no more Paradox of Continuity, no more 'language as a representational system', nothing. Now there's continuity and communication.

We are convinced that this will appall many readers. It is, after all, a rather radical change, indeed. As a matter of fact, we could articulate our review along this axis, and try to show that the continuist approach that Bickerton now embraces is hopeless, that no account of the origins of language will ever work as long as it is based on the idea that communication evolves. Because communication does not evolve, nor other functions for that matter.

But we are not going to do this. And the reason why we are not going to do it is this: Despite the fact that we have criticized these ideas elsewhere (Balari & Lorenzo 2009a, in press), we are coming to the conviction that this is becoming a sterile debate—that one either believes in the 'communicative approach' or one doesn't, that an intellectual wall has been built between two radically different ways of approaching the problem of the origins of language, and that very little can be done to permeate this barrier. We are still pretty convinced that the communicative approach is wrong and, therefore, our review of AT could easily dismissed as a matter of ideology, so we are not going to take this line of attack. Especially, because the problems with AT go well beyond this debate, and it is on these problems we want to focus our attention to.

So, what's *so* wrong with AT? We'll get to that in a moment, but first, a quick overview of the proposal. AT's main theme is that language could have evolved from *any* animal communication system should the appropriate conditions had been met. The crux of the matter is that such conditions are not internal to the organism, but, rather, external to it—they are intimately linked to the kind of ecological niche the organism in question entered (or constructed; more about this below) at some point during its life history. This is how the Paradox of Continuity is (dis)solved: Continuity between animal communication systems is possible, but it is not observed between primates and humans because extant primate species do not occupy the appropriate ecological niche, whereas the ancestors of humans did enter this niche. Language is the inevitable consequence of the kinds of selective pressures that certain niches impose over the communication systems of organisms. According to Bickerton such pressures are intimately connected to the kind of foraging strategy 'chosen' by the

organism, which, in the case of human's ancestors corresponds to what Bickerton calls 'high-end scavenging'.

The story goes more or les like this: After reviewing some circumstantial evidence, the author concludes that, at some point during the transition between Australopithecines and *Homo*, the old 'low-level scavenging' niche was abandoned and the construction of the new 'high-level scavenging' niche started. Low-level scavenging was the strategy our ancestors were forced to use in order to acquire food resources once a number of climatic changes turned a forest environment, rich in fruit, nuts, seeds and roots, into a dry savanna, where such resources are notably scarce. The only alternative was to profit from the carcasses of dead animals, but, since hyenas, vultures and other carnivores were tough competitors, our ancestors could only benefit from what these other animals left to them, namely bones. For a while, then, we fed on bone marrow, which we acquired thanks to very crude stone tools we used to break the thicker bones that carnivore teeth were unable to pierce. Suddenly, however, things change and there seems to be evidence that somehow our ancestors managed to outcompete other scavengers and were getting there first and keeping for them the best morsels of meat. How all this came to happen? The answer, of course, is that we invented language. Well, not really, just some kind of protolanguage, but already with a critical feature that was to pave the way to full-fledged language. The critical feature in question is displacement. Already in *Language & Species*, Bickerton (1990: 152–154) hinted at the idea that scavenging might have something to do with the origins of language, because it somehow seems to require recruiting the other members of the group to get the most of the food resource. In AT, he takes this idea to the extreme (there is some continuity in Bickerton's ideas after all) and he presents us with the following just-so story: Our ancestors *needed* language—protolanguage with displacement, that is—once they had started constructing the high-end scavenging niche. They needed it to tell the other members of the group about what they had found, that a nice deinotherium was lying there, dead stiff, over the hill, and thus, that they could all run there with their crude stone tools and throw them at their competitors, and with their cries freak them all out and frighten them all away in order to keep the whole carcass for themselves. So the story goes.

"But wait, *how* did language came to be?", you may be tempted to ask. Easy, Bickerton answers, it's just a matter of niche-construction. Look at ants, look at bees, they are also foragers who recruit. And they have displacement. It's just a matter of getting into the appropriate niche and it will generate the appropriate needs to start the appropriate feedback process eventually leading to language. It's just inevitable.

"But, sorry, wait again", you would probably object, "ants and bees *may* have displacement, but they *don't* have language, do they?". So there must be some other important difference between them and hominids. Big brains, perhaps? Not necessarily so, Bickerton retorts. Actually, it's the other way round, brains got bigger because of language. It's a fallacy to believe that one needs a big brain to get language.

So, the conclusion is clear: "Other animals didn't get language because, bottom line, they didn't need language." (AT: 24).

Now we have all the necessary elements of Bickerton's model in place, but, for the reasons exposed above, we would like to concentrate on just two of them. First, by appealing to the 'Big Brain Fallacy', Bickerton denies any role to proximate/internal causes in the origins of language, leaving all the work to ultimate/external causes, namely behavior. Second, the evolutionary mechanism responsible of the emergence of language was a process of niche-construction in the context of high-end scavenging which—necessarily—eventually yielded a communication system with the property of displacement.

In the two remaining sections of this review we would like to show that the first assumption is a vivid example of the most cavalier oversight of current knowledge concerning brain evolution ever; as for the second point, we will argue that Bickerton misinterprets for his own benefit Niche-Construction Theory (NCT; Odling-Smee *et al*. 2003) to turn it into a naive environmentalism of sorts, going well beyond the environmentalism exemplified by Marx and Engels, or, for that matter, Rousseau.

3.   The Big Brain Fallacy, or Why Ants and Bees Could Have Language (If They Don't Have It Already)

If you want to pass the buck to the environment in accounting for some evolutionary process, you'd better get rid of all possible proximate causes that might interfere in your goals. That's what the Big Brain Fallacy is for.

We'll never be able to express Bickerton's position better than him, so we'd rather quote him in full:

> If, as some have claimed, language was an invention of folk with big brains, it would be doubly unique. In addition to being the only system of its kind in nature, it would be the only biologically based behavior that had ever been consciously and deliberately created. And if you believe we can deliberately create biologically based behaviors, I have a couple of bridges you might contemplate purchasing.
>
>   But the real clincher is this. Brains don't grow by themselves, of their own volition; they grow because animals need more brain cells and connections to more effectively carry out any new things they are beginning to do. In other words, brain-size increase doesn't drive innovation—innovation drives brain-size increase.   (AT: 34)

Sic. No comment about the first paragraph. It speaks for itself. As for the second, we could start by saying that, of course, brains don't grow by themselves, of their own volition, but who *ever* said anything like that? (No references given concerning that particular point in AT.) This would really be a rather outrageous view of how organisms and organic structures develop and evolve. We're still wondering where Bickerton picked up this idea. Certainly not from Aristotle, nor from the preformationists, nor from the recapitulationists either; certainly not from any contemporary work on biology. Looks like a straw man.

The remaining lines of the second paragraph deserve some comment, however. They are important because they concern the directionality of the arrow of causality in developmental/evolutionary processes. As it is clear from the quote above, Bickerton sees a single causal arrow pointing in one direction only,

namely from the outside to the inside. In order to fully appreciate the kind of naive biological thought underlying this contention, we need to get into a detailed analysis of Bickerton's interpretation of niche-construction, a task that we undertake in the next section, but let us point out for the time being that such a view where all the power goes to the environment runs counter all current conceptions of how evolutionary processes work; it's not even Darwinian. It is not Darwinian, because, among other reasons, by attributing such an immense power to the environment, it is neglecting the effect of any existing structural constraint on organisms that might prevent them to evolve in some specific directions. For Bickerton, given some appropriate genetic fiddling, the appropriate environmental conditions could turn ants and bees (they have displacement, don't they?) into full-fledged linguistic beings (if they aren't already; see AT: 136–137).

Turning our attention back to brains, we would be much more cautious before making such a categorical statement concerning brain evolution. As pointed out by Striedter (2005: 134), "we are still far from consensus when it comes to the biological significance of evolutionary changes in overall brain size", so we do not know exactly why brains got bigger. What we *do* know, however, is that brains, since they were invented, have been getting steadily bigger and bigger (at least in vertebrates; Finlay & Darlington 1995, Striedter 2005); and we also know that some of these episodes of brain growth correlate with behavioral innovations (Striedter 2005), which is hardly surprising given the fact that all behavior supervenes on brain structure and organization. Granted, but from this it does not follow that novel behavior actually *causes* brain-growth. In fact, there seem to be reasons to expect that the opposite is true. These reasons are mostly developmental and concern the fact that developmental programs are extremely conservative (Shubin *et al*. 1997, 2009), and the case of brains is no exception (Finlay & Darlington 1995), so it is now well accepted that a very important source of novelty are some specific small perturbations of developmental programs that often give rise to some kind of heterochrony (Gould 1977, Parker & McKinney 1999). Again, this is probably not a knock-down argument against Bickerton's position, but it becomes one when coupled with the following quote:

> As for 'the rewiring of the brain,' brains don't rewire themselves for no reason, or because they just feel like a spot of rewiring. Brains rewire themselves, to the extent they do, because things are happening in the outside world, things that give individuals with rewired brains an advantage over individuals without them. (AT: 183)

Well, *this* is plainly false. Brains do rewire themselves when they get bigger: "[B]rains, like companies, must reorganize as they increase in size" (Striedter 2005: 127). And for very good reasons. As pointed out by Hofman (2001), as a brain grows bigger, soon a problem of connectivity arises given the fact that cortical and subcortical structures do not grow at the same rate (Finlay & Darlington 1995, Kaskan & Finlay 2001, Striedter 2005), which is linear for the latter but exponential for the former. As a consequence of that, and to ensure efficient neuronal communication, brains tend to compartmentalize and to develop laminae with groups of highly interconnected neurons (Deacon 2000), the net result being

that bigger automatically means rewired. Thus, not only do brains rewire them-selves, but when they do it, it is because something is happening in the *inside* world, not outside.

We can now see how the 'Big Brain Fallacy' collapses: Even if brain growth could eventually be attributed to some behavioral change (which is dubious), such increase in size would have necessarily implied a general rewiring and reor-ganization of the system in order to preserve its efficiency, with unpredictable consequences for the cognitive capacities of the organism and, no doubt, with an enormous potential for introducing behavioral novelty (Balari & Lorenzo 2008, 2009b). Thus, the more parsimonious (and logical) hypothesis given current evi-dence is that brain growth drives innovation, not that innovation drives brain-size.

4. How Bickerton Invented Niche-Construction

Now it's time to be fair. Fair with NCT, which cannot be held responsible of the fact that the evolutionary model adhered to by Bickerton in AT receives the same name.

Let us first try to disentangle the different threads of the fabric spun by the author in order to find the keys of the evolutionary mechanisms that he seems to have in mind. We'll start with a couple of quotes:

> But behavior and environment aren't watertight compartments—they're intimately linked; they shape each other into a lock-and-key fit. And fit is precisely what is meant by 'adaptation'. (AT: 64)

> So a feedback process begins, a two-way street in which the animal is developing the niche and the niche is developing the animal, until you get the lock-and-key fit between animal and niche that makes people say, "But there must be a designer!" (AT: 99–100)

At first sight, this looks like garden-variety adaptationism. We even find the very same metaphors Richard Lewontin once used to define it:

> The concept of adaptation implies that there is a preexistent form, problem, or ideal to which organisms are fitted by a dynamical process. The process is adaptation and the end result is the state of being adapted. Thus a key may be adapted to fit a lock by cutting and filing it. (Lewontin 1985a: 67)

Note that, despite of the fact that Bickerton appears to assume that orga-nisms posses some kind of causal power over environments, his characterization of the process is virtually identical to that of Lewontin's: The niche is there, the organism enters it, and a process begins which will eventually culminate when the perfect fit is attained, at which point it will stop ("until you get the lock-and-key fit"). That niches are there, waiting, until some organism decides it's about time to move and start making a new living in it follows from a number of statements we find in AT, like this one, for example:

> The size niche exists, permanently, within any order, simply because if you're bigger than anything else around, you're virtually invulnerable to attack. (AT: 122)

So far, then, we seem to be within the bounds of a Darwinism of the most orthodox kind, as the model doesn't look really different from those of classical sociobiology (Wilson 1975, Dawkins 1982), where genes—and only genes—are the one and only mechanism of variation, and where natural selection, that is, the selective pressures generated by the different environments (or niches), acts as the only causal power, capable of modifying allelic distributions within populations and favoring only those alleles correlating with those behaviors which happen to be apt for the niche in question. But, although Bickerton recognizes that the Extended Phenotype model of Dawkins (1982) comes close to what he has in mind with his niche-construction, he explicitly rejects it (AT: 99) to plunge headfirst into something entirely different.

Let us go back to the classical model for a while: Genes produce random variation and natural selection favors one or another variant depending on the conditions of the environment. That is, the internal mechanisms of variation (genes) express themselves into phenotypes, which are visible (external) for natural selection. Thus, causality follows an outside-inside direction, but note that a basic precondition for this kind of causality to work is that some variation pre-exists that is internal in nature: Classical Darwinism never paid too much attention to developmental processes, because it always assumed that the gene-phenotype relation is simple or negligible; but what Darwinism never denied is that variation has its origins in *internal* mechanisms. What is Bickerton's view on this particular issue? Two more quotes from AT:

> Most often than not, behavior changes first, then the genes change to keep
> pace with it.   (AT: 96)

> [O]nly external events can shape internal events, because only external
> events are visible to natural selection.   (AT: 185)

At first sight, this is not too different from the classical model: The arrow of causality travels from the outside to the inside. Note, however, a critical difference concerning the nature of variation (and, maybe, also inheritance), since Bickerton is explicitly inverting the directionality of the phenotype-gene relation, which explains that we find him making such preposterous claims as the following:

> It was, after all, the development of powered flight that eventually caused
> genes normally devoted to building front legs to express themselves, among
> birds and bats, in the form of wings.   (AT: 131)

So, first came flight and only thereafter genes started making wings. Note how Bickerton is here inserting a new causal level between the organism and the environment, a behavioral/functional level with an enormous creative power which, in fact, relegates genes (and any other proximate cause) to the role of mere puppets, ready to cater for whatever needs the organism may have: If we entered the flying niche, the need of flying would be created in us and we would start flying; later, genes will make wings for us: Like good old Arthur Dent in *The Hitchhicker's Guide to the Galaxy*, you just have to fail to hit the ground.

Beware, this is not to be confused with other models where an explicit

attempt has been made to integrate behavior into the causal chain of developmental processes in order to overcome the classical view of development as a simple, linear, gene-phenotype relation, and to see it instead as a complex process with multiple causal forces (e.g., Oyama 2000, Jablonka & Lamb 2005). Not at all. In AT, Bickerton explicitly despises the relevance of development in evolution (AT: 130): For him, development is in charge of building structures, parts of the body, not behaviors and, as he has explicitly stated several times in AT, evolutionary changes are driven by behaviors, so it's behaviors who determine what developmental processes have to do, not the other way round. Nothing in AT suggests the possibility of integrating everything that concerns development in a new notion of environment with behavior being one of these elements. As we will see presently, however, this integrative perspective *is* a constitutive idea of the notion of niche-construction.

This is not NCT; it is not classical Darwinism either, neither of the orthodox nor of the heterodox kind; it is, if anything, Lamarckism, but a crude and clumsy Lamarckism of sorts—typical of the most uninformed interpretations of the notion of the inheritance of the acquired characters; with, perhaps, a pinch of Rousseau. It is, in any case, a serious perversion of the ideas and theoretical underpinnings of NCT. Let's see why.

Since its inception, NCT has presented itself as a spin-off of the original proposals by Lewontin (1983, 1985b, 2000) to dispose of the adaptationist lock-and-key metaphor in favor of a constructive, dialectical model in which genes, organism and environment are an integral part of the same cyclic developmental process, where organism and environment co-construct each other in a never ending process (unless the organism goes extinct, of course). It is, in other words, an attempt to dissolve the traditional dichotomy between external and internal causes, and to grant the same role in the evolutionary process to all intervening factors, including, of course, natural selection (Laland *et al*. 2001a, Day *et al*. 2003), and thus recognizing the power of the organism to modify the selective pressures acting over it. Obviously, as it has been pointed out several times by the proponents of NCT (Laland *et al*. 1996, 2000, 2001b, 2003, and Laland & Brown 2006), these factors often pertain to the realm of what traditionally is seen as external, to the sphere of behavior or culture, but they never attributed to these factors the qualities Bickerton attributes to them in his own model.

"We [don't] have to struggle with the high-powered math that Odling-Smee and his colleagues use to justify their ways to population geneticists", Bickerton advises us (AT: 103). But if we do not follow his advice and take the trouble of examining their math, we will immediately discover that their models are a variety of epistatic selection population models with multiple *loci* (Font-devila & Moya 1999: chap. 6), where a third element R is added representing the environmental factor that is affected by the presence/absence of the different linked alleles, which, in turn, may affect the frequency in the population of one or another haplotype (e.g., Laland *et al*. 1996). That is, and following the two-locus model of Laland *et al*. (1996), we have two *loci* A and E, with alleles A/a and E/e, respectively, yielding the haplotypes AE, Ae, aE, and ae, such that variation in E determines changes that may occur in R, and variation in A determines the organism's fitness with respect to the environment modified by the phenotypic

traits defined by E. Thus, a feedback model is defined where variations in E cause changes in R which, in turn, favor one or another allele of A, dynamically bringing about changes in the frequencies of the different haplotypes in the population. We don't need to go into the fine mathematical details to see that this model has little, if anything, to do with the Bickertonian idea that "behavior changes first, then genes change to keep pace with it". Quite to the contrary, for the process to be able to start, a necessary precondition is the existence of some kind of genetic variation in the population. In a later paper (Laland *et al*. 2001b), the model was extended to the case where one of the two *loci* is not a genetic character, but some specific cultural activity, corresponding, for example to the oft-cited phenomenon of the expansion of the gene conferring greater adult lactose tolerance among certain human populations in association with persistent domestication of cattle and dairying activities (Durham 1991, Richerson & Boyd 2005). Clearly, in both models, as a consequence of some behavioral novelty, either with a genetic or with a cultural basis, the only genetic changes we observe concern the frequency of some previously existing allele in the population. What we certainly do not see is the kinds of changes Bickerton suggests in AT, where genes, with their plasticity and flexibility seem to be able to sanction any possible and imaginable behavior invented by the organism to satisfy its needs in the niche it currently occupies.

NCT is yet another attempt to integrate certain environmental and cultural factors in a more global and complete picture of evolutionary processes, with the aim of showing that the classical models represent a too simplistic view of evolution. It is therefore not surprising that several people had started thinking about a new evolutionary synthesis in biology, where the different alternative views are integrated that share the construction metaphor in their view of evolution, such as ecosystems ecology (Jones *et al*. 1994, 1997, and Wright & Jones 2006), evolutionary developmental biology (Hall 1999, Minelli 2009), phenotypic/developmental plasticity models (West-Eberhard 2003), ecological developmental biology (Gilbert 2001, Gilbert & Epel 2009) and, of course, NCT. In fact, some bridges have already been built between NCT and evo-devo, for example (Laland *et al*. 2008), and the shape that is starting to take form is the antithesis of the poor and uncouth biology that Bickerton is proposing. That's the difference between chatting about biology and actually doing some.

5.   Conclusion

As we advanced at the beginning of this review, our criticisms do not focus on AT's communicative and continuist approach. We decided to skip that. It's just one of the many secondary factors that make AT a bad book.

Indeed, we could've added that its style is arrogant, full of irrelevant personal anecdotes, florid just-so stories about our ancestors scavenging in the savanna, and the odd bad joke dropped here and there just for the fun of it; plus some distasteful ad hominem argument to top it all. Or we could've said that AT doesn't actually explain how language evolved. Some kind of protolanguage with displacement, maybe, but language, with such things as agreement, questions, anaphora, in a nutshell with all those features that make the structure

of natural languages so complex, not really. Just skip everything and go to the last chapter and watch Bickerton's hand waving.

But these are just ancillary matters, secondary issues on the basis of which one is not supposed to judge scientific writing. So we centered our attention to scientific questions, and it is our hope to have shown that AT is not just bad evolutionary linguistics, it's bad science, very bad science. If it is science at all. It's something not even able to match the pre-Darwinian environmentalism of Marx and Engels.

References

Balari, Sergio & Guillermo Lorenzo. 2008. Pere Alberch's developmental morphospaces and the evolution of cognition. *Biological Theory* 3, 297–304.

Balari, Sergio & Guillermo Lorenzo. 2009a. Comunicación: Donde la lingüística evolutiva se equivocó. Report de recerca CGT-09-10, Centre de Lingüística Teorica, Universitat Autònoma de Barcelona.

Balari, Sergio & Guillermo Lorenzo. 2009b. Computational phenotypes: Where the theory of computation meets evo-devo. *Biolinguistics* 3, 2–60.

Balari, Sergio & Guillermo Lorenzo. In press. ¿Para qué sirve un ballestrinque? Reflexiones sobre el funcionamiento de artefactos y organismos en un mundo sin funciones. *Teorema: Revista Internacional de Filosofía*.

Bickerton, Derek. 1990. *Language & Species*. Chicago: University of Chicago Press.

Dawkins, Richard. 1982. *The Extended Phenotype*. Oxford: Oxford University Press.

Day, Rachel L., Kevin N. Laland & John Odling-Smee. 2003. Rethinking adaptation. The niche-construction perspective. *Perspectives in Biology and Medicine* 46, 80–95.

Deacon, Terrence W. 2000. Heterochrony in brain evolution. Cellular versus morphological analyses. In Sue Taylor Parker, Jonas Langer & Michael L. McKinney (eds.), *Biology, Brains, and Behavior: The Evolution of Human Development*, 41–88. Santa Fe, NM & Oxford: School of American Research Press/James Currey.

Durham, William H. 1991. *Coevolution: Genes, Culture, and Human Diversity*. Stanford, CA: Stanford University Press.

Finlay, Barbara L. & Richard B. Darlington. 1995. Linked regularities in the development and evolution of mammalian Brains. *Science* 268, 1578–1584.

Fontdevila, Antoni & Andrés Moya. 1999. *Introducción a la genética de poblaciones*. Madrid: Síntesis.

Gilbert, Scott F. 2001. Ecological developmental biology: Developmental biology meets the real world. *Developmental Biology* 233, 1–12.

Gilbert, Scott F. & David Epel. 2009. *Ecological Developmental Biology. Integrating Epigenetics, Medicine, and Evolution*. Sunderland, MA: Sinauer.

Gould, Stephen J. 1977. *Ontogeny and Phylogeny*. Cambridge, MA: The Belknap Press.

Hall, Brian K. 1999. *Evolutionary Developmental Biology*, 2nd edn. Dordrecht: Kluwer.

Hofman, Michel A. 2001. Brain evolution in hominids: Are we at the end of the

road? In Dean Falk & Kathleen R. Gibson (eds.), *Evolutionary Anatomy of the Primate Cerebral Cortex*, 113–127. Cambridge: Cambridge University Press.

Jablonka, Eva & Marion J. Lamb. 2005. *Evolution in Four Dimensions. Genetic, Epigenetic, Behavioral, and Symbolic Variation in the History of Life*. Cambridge, MA: MIT Press.

Jones, Clive G., John H. Lawton & Moshe Shachak. 1994. Organisms as ecosystems engineers. *Oikos* 69, 373–386.

Jones, Clive G., John H. Lawton & Moshe Shachak. 1997. Positive and negative effects of organisms as physical ecosystem engineers. *Ecology* 78, 1946–1957.

Kaskan, Peter M. & Barbara L. Finlay. 2001. Encephalization and its developmental structure: How many ways can a brain get big? In Dean Falk & Kathleen R. Gibson (eds.), *Evolutionary Anatomy of the Primate Cerebral Cortex*, 14–19. Cambridge: Cambridge University Press.

Laland, Kevin N. & Gillian R. Brown. 2006. Niche construction, human behavior, and the Adaptive-lag Hypothesis. *Evolutionary Anthropology* 15, 95–104.

Laland, Kevin N., John Odling-Smee & Marcus W. Feldman. 1996. The evolutionary consequences of niche construction: A theoretical investigation using two-locus theory. *Journal of Evolutionary Biology* 9, 293–316.

Laland, Kevin N., John Odling-Smee & Marcus W. Feldman. 2000. Niche construction, biological evolution, and cultural change. *Behavioral and Brain Sciences* 23, 131–175.

Laland, Kevin N., John Odling-Smee & Marcus W. Feldman. 2001a. Niche construction, ecological inheritance, and cycles of contingency in evolution. In Susan Oyama, Paul E. Griffiths & Russell S. Gray (eds.), *Cycles of Contingency: Developmental Systems and Evolution*, 117–126. Cambridge, MA: MIT Press.

Laland, Kevin N., John Odling-Smee & Marcus W. Feldman. 2001b. Cultural niche construction and human evolution. *Journal of Evolutionary Biology* 14, 22–33.

Laland, Kevin N., John Odling-Smee & Scott F. Gilbert. 2008. EvoDevo and niche construction: Building bridges. *Journal of Experimental Zoology (Mol. Dev. Evol.)* 310B, 549–566.

Lewontin, Richard C. 1983. Gene, organism and environment. In D.S. Bendall (ed.), *Evolution: From Molecules to Men*, 273–285. Cambridge: Cambridge University Press.

Lewontin, Richard C. 1985a. Adaptation. In Richard C. Lewontin & Richard Levins, *The Dialectical Biologist*, 65–84. Cambridge, MA: Harvard University Press.

Lewontin, Richard C. 1985b. The organism as the subject and object of evolution. In Richard C. Lewontin & Richard Levins, *The Dialectical Biologist*, 85–106. Cambridge, MA: Harvard University Press.

Lewontin, Richard C. 2000. *The Triple Helix: Gene, Organism and Environment*. Cambridge, MA: Harvard University Press.

Minelli, Alessandro. 2009. *Forms of Becoming: The Evolutionary Biology of Development*. Princeton, NJ: Princeton University Press.

Parker, Sue Taylor & Michael L. McKinney. 1999. *Origins of Intelligence: The Evolution of Cognitive Development in Monkeys, Apes, and Humans*. Baltimore,

MD: The Johns Hopkins University Press.

Odling-Smee, John, Kevin N. Laland & Marcus W. Feldman. 2003. *Niche Construction: The Neglected Process in Evolution*. Princeton, NJ: Princeton University Press.

Oyama, Susan. 2000. *The Ontogeny of Information*, 2nd edn. Durham, NC: Duke University Press.

Richerson, Peter J. & Robert Boyd. 2005. *Not by Genes Alone: How Culture Transformed Human Evolution*. Chicago: University of Chicago Press.

Shubin, Neil, Cliff Tabin & Sean Carroll. 1997. Fossils, genes and the evolution of animal limbs. *Nature* 388, 639–648.

Shubin, Neil, Cliff Tabin & Sean Carroll. 2009. Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823.

Striedter, Georg F. 2005. *Principles of Brain Evolution*. Sunderland, MA: Sinauer.

West-Eberhard, Mary Jane. 2003. *Developmental Plasticiy and Evolution*. New York: Oxford University Press.

Wilson, Edward O. 1975. *Sociobiology: The New Synthesis*. Cambridge, MA: Belknap Press.

Wright, Justin P. & Clive G. Jones. 2006. The concept of organisms as ecosystems engineers ten years on: Progress, limitations, and challenges. *BioScience* 56, 203–209.

Sergio Balari
Universitat Autònoma de Barcelona
Department de Filologia Catalana
Facultat de Lletres, Edifici B
08193 Bellaterra (Barcelona)
Spain
sergi.balari@uab.cat

Guillermo Lorenzo
Universidad de Oviedo
Departmento de Filología Española
Campus El Milán
33011 Oviedo (Asturias)
Spain
glorenzo@uniovi.es

Response to Balari & Lorenzo

Derek Bickerton

Adam's Tongue, as its title surely indicates, is about the transition from the alingual state that characterizes all other species to something that might qualify as a genuine precursor of language, thereby opening the road for (*not*, *pace* Balari & Lorenzo—henceforth B&L—making "inevitable") the subsequent development of true language. It is thus illogical to complain, as B&L do, that it lacks explanations of "agreement, questions, anaphora" and similar components of full human language; inevitably, given the book's scope, such features of later development are touched on briefly, if at all. The book's secondary goal is to explain why human cognition as well as human communication differs so radically from that of other species.

B&L do not even try to deal with these issues. At least they are embarrassed by their omission of the first, and feel obliged to excuse it by claiming that the debate over the initial emergence of language is now "sterile" and too ideological to be pursued further (they fail to note the irony of proposing, in a biolinguistic journal, an embargo on the most crucial issue in biolinguistics). With regard to the second issue, they simply ignore it. Later I will suggest a possible motivation for this.

Adam's Tongue (Bickerton 2009a, sometimes abbreviated as AT, as used by B&L) is neither a textbook of evolutionary biology nor a primer of niche construction theory (NCT); it merely utilizes some concepts from these areas in its arguments. Yet amazingly it is on such incidental uses of biology and NCT that B&L almost exclusively focus. Granted, if I had really gone wrong here, the book's major contentions would be seriously flawed. But in fact it's B&L who are at fault. What they claim I say is far from what I actually say—sometimes even its exact opposite. And when they do quote me correctly, it turns out that I'm saying exactly what they're saying.

One claim they make is directly refuted by one of the sentence they themselves cite. In section 3, speaking of the directional arrow in evolutionary processes, they state "Bickerton sees a single causal arrow pointing in one direction only, namely from the outside to the inside" (p. 118). Just a couple of pages later, at the beginning of section 4, they quote my actual words (AT: 99–100): "So a feedback process begins, a two-way street in which the animal is developing the niche and the niche is developing the animal" (p. 120).

Thanks to the kindness of the editors, I have read the response by Balari & Lorenzo. Given that the latter fail to react to the substantive points I made, but instead continue to attribute to me views I have never held, I see no purpose in further comment on my part.

Other claims are simply bizarre, such as that according to me, "given some appropriate genetic fiddling, appropriate environmental conditions could turn ants and bees […] into fully-fledged linguistic beings (if they aren't already, see AT: 136–137)" (p. 119). In fact, all these two pages do is conclude, after a purely factual description of certain ant behaviors involving combinations of signals, that "this isn't quite like joining words. The shaking and the chemical trail may be meaningless in themselves, so it's more like the joining together of in-them-selves-meaningless speech sounds that we do to make words. But it's still concatenation of a kind; a kind that, primitive as it is, is found seldom if at all among the behaviors of other species".

Not a word, note, about the potentiality of ants to become "full-fledged linguistic beings", let alone their already being such. Indeed, if B&L had read a few pages further they would have seen that I explicitly reject any notion of ant or bee communication as even remotely ancestral to language: "How could things like these possibly be precursors of language? They're not […]. The question is not so much why our species got [language] and no other did, it's why *any* species got it at all" [AT: 144; original emphasis—DB]. So much for my alleged belief that "some appropriate genetic fiddling" could equip ants or any other species with language.

B&L also employ the Damning Context-Free Quote. Example: "It was, after all, the development of powered flight that eventually caused genes normally devoted to building front legs to express themselves, among birds and bats, in the form of wings" (AT: 131). Nothing in B&L's (p. 122) deconstruction of this piece ("So, first came flight and only thereafter genes started making wings") suggests that this sentence occurs, not as part of any general description of development, but in parentheses, in the middle of a paragraph devoted to an entirely different topic, where its sole function is to remind readers of a connection with two other topics, deep homology and flight, that have already been fully discussed elsewhere (AT: 129–130 and 9, respectively). In those discussions, readers will find all the details of gradualness and reciprocal influence in development that B&L accuse me of misunderstanding or ignoring.

On other occasions, B&L state my views, pour scorn on them, then express similar views themselves. This strategy is so remarkable that it demands at least two specific examples.

Example 1: They respond to my claim that behavioral innovations drive brain re-wiring, rather than vice versa, with "Well, *this* is plainly false. Brains do rewire themselves when they get bigger". And what makes them get bigger? B&L (p. 120): "[S]ome of these episodes of brain growth correlate with behavioral innovations".

Example 2: They describe my treatment of NCT as "a serious perversion" of that theory. But here, side by side, are our respective summaries of NCT (see if you can spot the difference):

> [A] constructive, dialectical model in which genes, organism and environ-
> ment are an integral part of the same cyclical developmental process, where
> organism and environment co-construct each other in a never-ending
> process.   (B&L: 124)

> [A]nimals themselves modify the environments they live in, and these modified environments, in turn, select for further genetic variations in the animal. So a feedback process begins, a two-way street in which the animal is developing the niche and the niche is developing the animal.   (AT: 99)

The only difference I can see is that mine is fourteen words longer but a whole lot easier to understand. What's almost unbelievable is that B&L themselves have already cited the second half of it, not as a summary of NCT, but with the comment "this looks like garden-variety adaptationism" (p. 121). How can they accuse me of perverting a theory if they can't even recognize that theory when they see it?

Heinous as they are, B&L's sins of commission pale before their sins of omission. I have already referred to their refusal to address one of the two core issues of the book—how language got off the ground (to add insult to injury, in their conclusion they dismiss this issue as merely "one of the many secondary factors" (p.124), though it occupies the first two-thirds of the book and should be central to any discussion of language evolution.) I shall therefore confine myself to the second issue, the role of language in forming human cognition.

I argue that, instead of enhanced cognition giving rise to language, as so may believe, language (from its very earliest manifestations) gave rise to enhanced cognition. In addition, I propose specific mechanisms by which this might have come about. Those proposals may be wildly wrong, but you wouldn't learn that from B&L's review. Nor would you know that the book contains extensive criticism of a currently popular approach to language origins that places those origins in the mind rather than in communication (Hauser *et al*. 2002)—an approach to which, though they are somewhat coy about it, B&L obviously subscribe. Rather than answer these criticisms (to which one and a half out of a dozen chapters are devoted) B&L prefer to pretend that they don't exist. And I suspect I know why.

The non-communicative approach holds the following beliefs (fully documented in *Adam's Tongue*):

(A)   The key development in language, perhaps the only one unique to it, was recursion, created by "Merge".

(B)   Before language was "externalized", Merge created recursive structures in the mind, linking concepts.

(C)   The concepts that Merge linked had to differ from animal concepts in that they did not, like the latter, refer to "mind-independent entities" (i.e. direct-ly, to real-world objects), but instead were symbolic in nature, representing abstract categories (Chomsky 2010).

In other words, before recursion could operate (therefore, before language could start), a new type of concept had to emerge. Where did such concepts come from? How and why did they form? In *Adam's Tongue* I try to answer such questions. My answers may not be correct, but they are at least answers.

The non-communicative approach hasn't even got that far. It has no answers. Supporters of this approach cannot afford even to look at the language-cognition interface, since doing so would force them to admit this gaping hole in their theory, and probably also to concede that, contra (A), there may be at least two components to the Faculty of Language (Narrow)—symbolic concepts as well as recursion. But, as *Adam's Tongue* also shows, recursion itself is dubious as a uniquely language-devoted function. Recursion can be defined in two ways; basic assumptions of Minimalism show that the stronger is simply an artifact of earlier generative formulations, while the weaker would hold for a wide variety of non-linguistic behaviors by both humans and other animals (see also Bickerton 2009b). But again, not a word about any of this from B&L.

The motives behind this review should by now be apparent; likewise its choice of targets. B&L are doubtless familiar with the nest-defense strategies of golden plovers (Byrkjedal 1989). When its eggs or nestlings are at risk from an approaching predator, this plover tries to draw the predator away by elaborate, albeit bogus behaviors (feigning wing injury, limping along on the ground, etc.), B&L's review has a similar distracting function. Finding their cherished ideas menaced by the central arguments of *Adam's Tongue*, they avoid any confrontation with those arguments (just as the plover avoids any confrontation with the predator) and instead seek to draw attention away from the vulnerable target (just as the plover does) by moving the discourse elsewhere, even though, in order to do so, they have to resort to behaviors as deceptive as the plover's dragging wing.

Ironically, B&L's negative review serves to support the core arguments of *Adam's Tongue* more convincingly than the most favorable review could have done. The latter might merely reflect the reviewer's bias. But B&L's review can lead to only one conclusion. The book's central claims are that language emerged from certain specific communicative uses, and that language created human cognition, rather than vice versa. If critics as determinedly hostile as B&L did not even try to refute these claims, it can only be because they were unable to do so.

References

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

Bickerton, Derek. 2009b. Syntax for non-syntacticians: A brief primer. In Derek Bickerton & Eors Szathmary (eds.), *Biological Foundations and Origins of Syntax*, 3–17. Cambridge, MA: MIT Press.

Byrkjedal, Ingvar. 1989. Nest defense behaviors of Lesser Golden Plovers. *Wilson Bulletin*, 101, 579–590.

Chomsky, Noam. 2010. Some simple evo-devo theses: How true might they be for language? In Richard Larson, Viviane Deprez & Hiroko Yamakido (eds.), *The Evolution of Human Language: Biolinguistic Perspectives*, 45–62. Cambridge: Cambridge University Press.

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

Derek Bickerton
University of Hawai'i Mānoa
Department of Linguistics
569 Moore Hall
1890 East-West Road
Honolulu, HI 96822
USA
derbick@hawaii.rr.com

Incidental Biology:
A Reply to Derek Bickerton's Response

Sergio Balari & Guillermo Lorenzo

When we first undertook the task of reviewing Derek Bickerton's (2009) *Adam's Tongue* (AT), we had to make a decision about what points we should focus our attention on. Bickerton, in his reply, finds it amazing that we almost exclusively focus on "incidental uses of biology and NCT" (p. 128). Indeed, we eventually decided against articulating our criticisms along the communicative vs. non-communicative debate for the reasons we already expressed in the review (henceforth SM), but also on the (implicit) assumption that any account of language origins should be based on a sound biological framework, that biolinguistics is about taking biology seriously, and that, should any attempt at an explanation of the origins of language be based on a wrong or inadequate use of biological notions, it would immediately be flawed, independently of how the communicative vs. non-communicative debate is eventually resolved. It appears that, at least as far as this point is concerned, Bickerton agrees with us: "[I]f I had really gone wrong here, the book's major contentions would be seriously flawed" (p. 128) And we think they are. That's the reason why we concentrated on them. But also because we do not think there are any biologically incidental issues in biology; nor in biolinguistics, for that matter.

AT is not "a textbook of evolutionary biology nor a primer in NCT" (p. 128) —true, but this is no excuse for making a sloppy use of these disciplines. In fact, with our selection of quotes we tried to show that, in spite of the fact that all the appropriate leitmotivs of NCT appear in the book, what Bickerton is actually advocating for in AT is a necessity scenario not too different from the one suggested by Marx and Engels more that 150 years ago. The main features of this necessity scenario are: (i) ecological niches preexist the organisms that enter them; (ii) the environment is sufficiently structured to create the need for a new character; (iii) if the organism in the niche is not constrained by its biology (more about this below) it will respond by developing the (behavioral) character that satisfies the need imposed by the environment; (iv) a feedback process ensues,

not too different from reinforcement in behaviorist psychology, that eventually fixes the character; and (v) in this feedback process genes and environment interact and this interaction is mediated by behavior.

From this follows that *any* organism could develop language should it enter the appropriate niche *and* should it have the appropriate biology. This last phrase, which is ours, not Bickerton's, essentially means that the organism must not be variation-limited (Számadó & Szathmáry 2006), i.e. that its biology must not include constraints that would make this change impossible. This probably is, by the way, what, according to Bickerton, is supposed to explain why ants and bees don't have language, although, without knowing exactly what these biological constraints are, we cannot rule out the possibility, given that they seem to occupy the appropriate niche, that they will eventually develop it.

Now, this is not NCT—we reaffirm our contention. It is, as we pointed out in SM, at its best, a variety of adaptationism; see our quote by Lewontin in SM, but also the following:

> Whereas classical evolutionary theory sees the organism as the key that has to fit into the environment's lock, both ecological developmental biology and niche construction see interactions between them. Niche construction emphasizes the ability of the organism to alter its environment; eco-devo emphasizes the ability of the environment to alter the developing organism.
> (Laland *et al*. 2008: 550).

At its worst, with the introduction of 'necessity' in whole picture, it is a modern version of Marxian or Rousseaunian environmentalism:

> [L]'usage de la parole s'établit ou se perfectionne insensiblement dans le sein de chaque famille, et l'on peut conjecturer encore comment diverses causes particulières purent étendre le langage, et en accélerer le progrès en le rendant plus necéssaire. Des grandes inondations ou des tremblements de terre environnèrent d'eaux ou de précipices des cantons habités; de révolutions du globe detachèrent et coupèrent en îles des portions du continent. On conçoit qu'entre des hommes ainsi rapprochés, et forcés de vivre ensemble, il dût se former un idiome commun plutôt qu'entre ceux qui erroient librement dans les forêts de la terre ferme.   (Rousseau 1755: 109)

The reason, again as we pointed out in SM, is causation. Thus, even if Bickerton appears to be willing to accept that development plays *some* role in evolutionary processes, he explicitly declares his skepticism (e.g. AT: 130) for evo–devo explanations, at least in the case of language origins. Indeed, given the way he conceives of his model, developmental processes are not regarded as independent causes, but rather, by appealing to the variation-limited vs. selection-limited dichotomy (Számadó & Szathmáry 2006), he is implicitly accepting that the ability of these processes to constrain is fully explained by natural selection, that is they are just the proximate manifestation of some ultimate, exogenous, cause, namely natural selection and/or behavior.

Compare now this picture with the following:

> An equally tenable approach, which we advocate, is to adopt 'reciprocal causation' in evolutionary explanations, in which characteristics of orga-

nisms are regarded as caused by interacting bouts of selection and con-
struction […]. One important ramification of reciprocal causation is that it is
philosophically sound to argue that developmental processes can be evolu-
tionarily causal, as they are not regarded as fully caused by earlier selection
on genes.   (Laland *et al*. 2008: 552)

So much for our sins of commision.

As for our sins of omission, it was not our intention to use SM as a means
to defend our own views, since we have already done that, in print, elsewhere
(Balari & Lorenzo 2008, 2009a, 2009b). However, since Bickerton accuses us of
using SM as some kind of diversion strategy, we would like to say something in
this connection. And we will start by saying that, no, we are not crypto-
Chomskyans garrisoned in the Hauser-Chomsky-Fitch barracks. Quite to the
contrary. We have raised a number of criticisms to the HCF paper (Hauser *et al*.
2002); thus, for example, in Balari & Lorenzo (2008: 6) we concluded that:

[O]ur proposal is compatible […] with the idea that the computational re-
gime subserving the human language faculty also subserves other, non-
linguistic aspects of human mentality, and, consequently that the narrow
faculty of language of [HCF] is, in fact, not specific to language.

And in Balari & Lorenzo (2009b: 45) we wrote:

We only slightly qualified [HCF's] assertion, pointing at evidence which
may indicate the presence of computational systems with degrees of com-
plexity similar to the FL but associated with cognitive capacities that give
rise to such disparate behaviors as nest building.

To which we could add (see Balari & Lorenzo 2009a) that HCF is often read and
interpreted as a defense of the specifically human and specifically linguistic
character of FLN, the computational aspect of language. This interpretation, we
believe, is correct; but only partly correct. HCF also, and perhaps above all, is a
defense of the singularity of language *qua* communication system and, in this
sense, a particular application of the research program on the evolution of the
communicative function developed by Hauser (1997), whose explicitly declared
main goal is to determine the causes underlying variation in natural
communication systems, human linguistic communication included. Indeed, one
should read as a direct appeal to this framework the fact that one of the basic
assumptions of HFC is that, despite the conspicuous discontinuity among the
systems on which communication is based in different species, these systems are
nothing else but several instantiations of a unique but highly diversified organic
function (2002: 1569; especially, fig. 1). HCF constitutes, then, no more and no
less than an attempt to provide an answer for the particular case of linguistic
communication within a broader research agenda seeking the causal factors
responsible for the piecemeal diversification of the communicative function in its
evolution within different species (Hauser 1997: 1–2).

Now, Bickerton attributes to us the opinion that "the debate over the initial
emergence of language is now 'sterile' and too ideological to be pursued further"
(p. 128) but our opinion, clearly expressed in SM, is that the discussion about

whether language has its origins in some form of ancestral communication or not is sterile and too ideological, which is quite different. To extend a bit this idea, given that we have been given to opportunity to do so (but, again, see Balari & Lorenzo 2009a), we contend that language does not have its origins in some form of ancestral communication; it *cannot* have its origins in *any* form of ancestral communication. The reason is, ultimately, conceptual: There is nothing in nature corresponding to what we ordinarily call 'communication'. Pheromone transmission in ants, the feigned wing injuries of golden plovers, the alarm calls of vervet monkeys, and human language are such an heterogeneous collection of phenomena that cannot in any sense be considered a 'natural class'. In a nutshell, our point of view is that 'communication' does not refer to a natural class that can be the object of evolutionary process (see also Klopfer 1973). Bickerton (1990) could have delivered evolutionary linguistics from this 'communicative fallacy' —directly instantiated in research programs like that of Hauser (1997), and indirectly, in that of HCF—but the truth is that, twenty years after, not only the fallacy still pervades this discipline, but also Bickerton has now become one of its most vehement advocates.

Finally, Bickerton reproaches us of not having addressed in SM what, in his opinion, is one the major contributions of AT: The explanation why the human mind is so radically different from other animal minds. He contends that it is language that created it and not the reverse, that is, he contends that the evolution of the human mind could not have given rise to something like language. True, we did not touch on this issue. It would not have been appropriate, since this is not really one of the topics developed in the book. As correctly pointed out by Bickerton in his response, "[AT] is about the transition from the alingual state that characterizes all other species to something that might qualify as a genuine precursor of language, thereby opening the road for […] the subsequent development of true language" (p. 128)—so AT's evolutionary perspective is exhausted with what Bickerton calls a "genuine precursor of language" and it does not make any genuinely new contribution concerning the role of 'true language' in the evolutionary modeling of the 'true human mind'.

Our suggestion to the interested reader, in order to compensate for AT's limitations, is to read Bickerton (1990) and Bickerton (1995).

References

Balari, Sergio & Guillermo Lorenzo. 2008. Pere Alberch's developmental morphospaces and the evolution of cognition. *Biological Theory* 3, 297–304.

Balari, Sergio & Guillermo Lorenzo. 2009a. Comunicación: Donde la lingüística evolutiva se equivocó. Report de recerca CGT-09-10, Centre de Lingüística Teorica, Universitat Autònoma de Barcelona.

Balari, Sergio & Guillermo Lorenzo. 2009b. Computational phenotypes: Where the theory of computation meets evo–devo. *Biolinguistics* 3, 2–60.

Bickerton, Derek. 1990. *Language & Species*. Chicago: University of Chicago Press.

Bickerton, Derek. 1995. *Language and Human Behavior*. Seattle: University of Washington Press.

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

Hauser, Marc D. 1997. *The Evolution of Communication*. Cambridge, MA: MIT Press.

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

Klopfer, Peter H. 1973. Does behavior evolve? *Annals of the New York Academy of Sciences* 223, 113–125.

Laland, Kevin N., John Odling-Smee & Scott F. Gilbert. 2008. EvoDevo and niche construction: Building bridges. *Journal of Experimental Zoology (Mol. Dev. Evol.)* 310B, 549–566.

Rousseau, Jean-Jacques. 1755. *Discours sur l'origine et les fondemens de l'inégalité parmi les hommes*. Amsterdam: Marc Michel Rey.

Számadó, Szablocs & Eörs Szathmáry. 2006. Selective scenarios for the emergence of natural language. *Trends in Ecology and Evolution* 21, 551–561.

Sergio Balari
Universitat Autònoma de Barcelona
Department de Filologia Catalana
Facultat de Lletres, Edifici B
08193 Bellaterra (Barcelona)
Spain
sergi.balari@uab.cat

Guillermo Lorenzo
Universidad de Oviedo
Departmento de Filología Española
Campus El Milán
33011 Oviedo (Asturias)
Spain
glorenzo@uniovi.es

Where is the Conflict between Internalism and Externalism? A Reply to Lohndal & Narita (2009)

Daniel Lassiter

In a recent issue of *Biolinguistics*, Lohndal & Narita (2009, henceforth L&N) reply skeptically to my attempt to sketch a theory that takes both internalist and externalist inquiry seriously (Lassiter 2008). My reply serves two purposes. First, L&N seriously misrepresent my work, and a certain amount of correction is needed. In particular, they seem to be laboring under the misconception that my proposal was intended as an argument *against* internalism, rather than an argument for a science of language that makes use of both internalist and externalist modes of inquiry and attempts to relate them systematically. A more important issue, however, is the evaluation of L&N's positive claims. I argue that their highly restrictive vision of the methodology of linguistics is dubious in light of the current landscape of the field, and potentially harmful to the development of a proper biolinguistics. I identify three separate notions of internalism that are lurking in L&N and argue that two of them are harmless but relatively uninteresting, while the third is both unmotivated and dangerously parochial. I conclude with some speculations about the place of 'externalist' inquiry in the emerging interdisciplinary field of biolinguistics.

1.   Corrections

Let me begin by emphasizing that, although I focus here on weaknesses in their account, I do think L&N make a real contribution to the discussion. In particular, they argue that my characterization of internalism is not faithful to Chomsky, and they offer an alternative characterization. This may well be right, and I will return to the characterization of internalism in section 2. Further, they state a widely held, but rarely articulated, set of assumptions with remarkable clarity, and this makes the task of evaluating these assumptions easier. With these positive aspects in mind, I will proceed here to what I see as flaws in their portrayal of my own position, and then to flaws in their own.

There are numerous inaccuracies in L&N's rendition of Lassiter (2008), indeed more than I can respond to in this space. I can only ask the reader, rather than judging my contribution from L&N's selective and distorted summary, to refer to the original. However, two important points of clarification are in order, without which it will be impossible to understand the following discussion. First, L&N claim repeatedly that my paper advocates "Dummett-type externalism"—a serious charge in some quarters. In fact I expended considerable energy to refute precisely this type of externalism, using Dummett as the prototype of an externalist whose theory is unworkable (Lassiter 2008: 611-617). The type of 'externalism' that I argued for in the remainder of the essay has little in common with Dummett except that it takes seriously the *intuitions* on which he bases his theory, rather than dismissing them out of hand. My theory was no more 'Dummett-type' than any theory whose practitioners make use of grammaticality intuitions is 'Chomsky-type'.

Secondly, L&N take my paper as a "criticism of the internalist project" and as an attempt to "suggest an alternative to the internalist science of language" (L&N: 330). This was not even remotely my intention, nor is this reading supported by my words (except under a certain extremely narrow definition of 'internalism', cf. section 2.3 below). It is true that my discussion did not *pre-suppose* internalism: This would have been problematic for my goal of convincing externalist-inclined philosophers that they can coexist with internalists. But I did not argue against internalism —far from it, my theory depended crucially on the fact that speakers possess internal linguistic representations that play a causal role in their linguistic productions, as I repeatedly emphasized.

The thrust of the paper was that the obvious difficulties in accommodating semantic externalism within linguistics—for example, problems about dialect continua identified by Chomsky (1986, 2000), and the usual dismissal of mentalism by externalists—can be resolved by careful attention to social and sociolinguistic details. The negative claim (not original to me) was that there are linguistic facts which cannot be explained solely by reference to the internal states of individuals; the main positive claim was that these same facts *can* be explained by reference to individuals' internal states and the way that individuals interact with each other. This is obviously not the same as saying that there are no internalistic facts about language, or that internalist inquiry is not an important, indeed crucial, aspect of the study of human language. My paper claimed that semantic externalism and internalism are, contrary to appearances, compatible and in fact complementary. This is not an argument *against* internalism under any reasonable construal of this term.

L&N's critique consists of at least three strands. One is a rather general defense of internalist inquiry which, since I am not an opponent of this methodology, misses the mark entirely. The second strand is a handful of specific criticisms of the proposal in Lassiter (2008), which I will not answer in detail because they are either beside the point (e.g., criticizing me for not defining 'accommodation' when that term is clearly defined and discussed in detail in references cited) or they do not seriously engage the details of the proposal under evaluation and its philosophical back-

ground, so that a reply would be little more than a rehash.

The third and most important type of criticism L&N raise is a generic indict-ment of externalist inquiry. Their argument here is interesting, to my mind, even though they do not even attempt to engage the large philosophical literature on semantic externalism that motivates Lassiter (2008), nor to discuss the specific questions work that that paper responds to (e.g., issues in the semantics of names and definite descriptions, or detailed questions about how semantic-externalist intu-itions interact with specific patterns of social interaction). Instead L&N focus on the question of whether externalist inquiry in general (N.B.: not just *semantic externalism*, the main concern of Lassiter 2008) is worth pursuing.[1] This is a rather blunt-edged argument, so it is not clear why they chose my paper, of all the huge body of rele-vant work in philosophy, linguistics, psychology, anthropology, etc., as a target for this criticism. However, their discussion raises a number of interesting questions which are worth responding to on their own merit. This is the issue which was hinted at in the previous paragraph, and which forms the core of L&N's paper: What is a reasonable construal of internalism, and in what sense is it in conflict with exter-nalist inquiry?

2.   Three Types of Internalism in L&N

2.1.   *Internalism by Definition*

The first kind of internalism lurking in L&N's discussion is internalism by definition: We simply define 'language' as 'internalist aspects of language', thus deriving the conclusion that externalist aspects are not 'language'. This idea seems to be present, for example, in the following quote:

> Lassiter claims that we can overcome such difficulties […] by incorporating
> some sociolinguistic notions into linguistic theory.   (L&N: 321)

Many linguists would have thought that sociolinguistic notions are already present in linguistic theory. (As it happens, sociolinguists do not appreciate being told that what they do is not 'linguistics' or that it is not 'theory'.) But the terminological question is not very interesting to my mind: If someone wants to define the social aspect of language use as part of the study of social psychology, for instance, nothing particularly interesting hinges on this choice.

---

[1]   Note that here and throughout I am using the unmodified term 'externalism' in the very broad sense that L&N do, as a description of inquiry into language use in any fashion. This is *not* what philosophers usually mean by 'externalism', and it is not how the term was used in Lassiter (2008). The latter notion—which I will call 'semantic externalism'—picks out a somewhat heterogeneous class of claims about the relationship between the meanings of words as an individual speaker uses them, and the usage of others with whom she is in contact. (This is intentionally vague, since there are many conflicting implementations of this idea which have little else in common.) Semantic externalism characterizes only a small part of the wide range of inquiry that counts as 'externalist' in L&N's sense.

2.2. *Internalism as Methodology*

I take it that L&N consider this to be their most important argument for internalism, since it contributes the title of their paper. This is puzzling: If the choice to ignore externalist aspects of language is simply a methodological one, then there should be no difficulty in allowing that other researchers, choosing a different methodology, may go about their business, and perhaps even uncover results that complement one's own. Two variants of this position appear in L&N. First:

> Chomskyan internalism […] just amounts to "the *methodological decision* […] to study *less*, prior to studying *more*".   (L&N: 329, citing Hinzen 2006: 161)

Can this really be all that internalism amounts to? If so, it is harmless but theoretically impotent. You decide to study less, hoping to get deeper; I decide to study more, hoping to get a broader view. Eventually, we may converge. In any case, it makes no sense to argue about which methodology is right. For example, you might study cellular biology without taking an interest in the evolutionary history of the relevant organism. I might study evolution, either in isolation or along with cellular biology. It would be strange indeed for one of us to argue that the other has chosen the wrong methodology.

On the other hand, "internalism as methodology" might mean this:

> We need a more complete understanding of the internal properties of I-language before we can even attempt to try to understand how individuals utilize them to deal with all sorts of E-language phenomena.   (L&N: 328)

Perhaps. But the claim that L&N are trying to make needs more justification than their plea that their own field (syntax, I take it) is not sufficiently well understood. First, a good deal has been learned about language use in sociolinguistics and pragmatics that is independent of the specific grammatical theory employed, undercutting the claim that a complete understanding of I-language must come first. But, more to the point, it's simply not true in general that explanatorily more basic areas must be fully understood before profitable research in higher-level fields can be undertaken. Darwin did a great deal of extremely important work on evolutionary theory without having any idea what the mechanisms of descent with modification were. Likewise, no one would seriously argue that all research in ecology should be put on hold until every detail of the structure and function of mitochondria is understood. As far as I can see, neither of L&N's arguments for "internalism as methodology" convey any more than their personal preference for investigating facts that can be given an account in strictly internalist terms. This is all well and good, but hardly justifies the dismissive attitude toward externalist inquiry that the authors display. Something more must be at play.

2.3. *"A Truly Scientific Explanation"*

In the end, L&N make only one argument that is genuinely in conflict with the

approach outlined in Lassiter (2008): The argument that externalist inquiry, in general, cannot be given a scientific basis. Since Lassiter (2008) is an example of externalist inquiry, it follows that Lassiter (2008) is not scientific. For instance:

> What internalists doubt is rather the feasibility of providing a serious science of any mind-external phenomena such as [normativity and communicative success].
> (L&N: 326)

> [G]eneral issues of intentionality, including those of language use, cannot reasonably be assumed to fall within naturalistic inquiry.
> (Chomsky 1995: 27, cited by L&N: 326)

> [I]nternalists never deny that there are complicated social aspects in the domain of language use; they just decide not to let these *unexplainable* aspects of language use enter into their naturalistic theory at the present stage of inquiry.
> (L&N: 330; emphasis added)

These are sweeping claims. If they are correct, a great number of people who take themselves to be doing research on human language and related topics have, in reality, been wasting their time. One would expect that such claims would be accompanied, say, by a thorough discussion of the large body of existing work on language use (see below) and an explanation of this dismissal making use of some clear-cut criteria for what counts as being 'scientific'. Instead, L&N make this assertion with virtually no argumentation except for a few references to Chomsky's philosophical works (Chomsky 1995, 2000), and a footnote quoting a definition of "serious science" from McGilvray (2002) with no indication of why it should be accepted, or its relevance to the specific issues at hand.

Frankly, I have a hard time seeing how this argument can be taken seriously. I know of no other discipline where scholars can dismiss each others' work on purely aprioristic grounds and others will accept this move as 'scientific'. Other examples of this habit are Chomsky's (2000) insistence that a theory of language use is impossible because it would have to be a "theory of everything"—never mind that numerous well-developed theoretical approaches to language use were already in existence when this pronouncement was made—or Narita's (forthcoming) claim that externalist inquiry is invalid because it disagrees with Descartes' notion of free will. To put it simply: In a "serious science" you cannot dismiss others' work because it disagrees with your philosophical preconceptions—or at least, you cannot do this and expect anyone to listen. Criticism of others' work will consist in a demonstration of some actual problem with it—for example, that it makes incorrect predictions about some domain—and not simply the fact that one does not feel comfortable with its assumptions. (Of course anyone can feel however they want about anyone else's work; they just can't expect other scholars to take this seriously as an argument.)

Further, L&N's argument—particularly in the third quote cited above—is uncannily similar to what Dawkins (1986) describes as the "argument from personal incredulity". As Dawkins observes:

> Even if the foremost authority in the world can't explain some remarkable biolo-
> gical phenomenon, this doesn't mean that it is inexplicable. Plenty of mysteries
> have lasted for centuries and finally yielded to explanation.   (Dawkins 1986: 39)

L&N would do well, I think, to be more cautious about what they describe as
"unexplainable". In particular, when evaluating a specific proposal (such as Lassiter
2008, or any of the references cited below), it is not enough to declare that they find
the phenomena in question mysterious, or that they have already designated that
domain as "unexplainable" or "unscientific" and that the proposal is thus wrong.
This is bad argumentation, and totally unconvincing; it also undermines the preten-
tions of being "scientific" that the authors claim to hold dear.

Even worse, there already exists an enormous body of work investigating and
theorizing about language use and closely related topics. I do not know whether
L&N are unaware of this work, or if they have some reason to dismiss it; but it is
irresponsible to make pronouncements like those quoted above without even consi-
dering it. I have in mind:

- Gricean and Relevance-Theoretic pragmatics (Sperber & Wilson 1986, Grice 1989, Horn 1989, and much other work)
- decision-theoretic and game-theoretic pragmatics (e.g., Parikh 2001, van Rooij 2003, Benz *et al.* 2006)
- experimental, theoretical, and empirical sociolinguistic work demonstrating speakers' sensitivity to social factors in their language use (e.g. Payne 1980, Le Page & Tabouret-Keller 1985, Labov 1994, 2001, Eckert & Rickford 2001)
- social-psychological work on accommodation, joint action, and related topics (e.g., Giles & Powesland 1975, Giles & Robinson 1990, Clark 1992, 1996)
- work on coordination and social norms in philosophy and economics (Lewis 1969, Skyrms 1996, Bicchieri *et al.* 1997, Young 1998, Bicchieri 2006)
- computational simulations of complex societies, including linguistic communi- cation (Epstein & Axtell 1996, Skyrms 2004, Epstein 2007)

This list could be much longer, but this should give a sense of how rich, varied, and
theoretically sophisticated the study of language use already is and promises to
become in the future.[2] (Sadly there is no room here for a detailed discussion of the
results and prospects of this large body of research; but the works cited speak for

---

[2]   Truth-conditional semantics has been the target of a certain amount of abuse from internalists,
and perhaps it belongs here as well. I omit it because I do not think that it is necessarily an
externalist enterprise. Nor, in my experience, do many of its practitioners, particularly the ones
who work in linguistics departments. Notions such as truth in a model are essentially mathe-
matical notions which do not care whether they are used for internalist or externalist purposes.
But if I am wrong, and it is true that semantics is essentially externalist, so much the worse for
L&N: They now have the burden of explaining for yet another field how so many researchers
have discovered so many systematic facts and compelling explanations in an area which
supposedly is not "serious science" or "naturalistic inquiry".

themselves.) It seems strange indeed, in this light, to hold to a speculative notion of what is or is not possible in the study of language use. Unless L&N are willing to dismiss all of this work as "unscientific", they will have to reconsider their position. Even if they are willing to do this, I see no reason to take such a position seriously, unless they can demonstrate point by point that all of this research is misguided. By any reasonable definition, it seems to me, there are numerous approaches to the study of language use that have claim to the title of "serious science".

3.   The Place of Externalist Inquiry within Biolinguistics

When I began my research into issues about internalism and externalism a few years ago, I knew of externalism only through Chomsky's writings, and I wanted to show once and for all that there is no place for semantic externalism in a science of language. As I read the classic works in the philosophical literature, however, I realized that semantic externalism was less a theory than a bundle of intuitions about how people would use words in certain counterfactual situations. Even though existing attempts to explain these intuitions were clearly unworkable, it did not follow that no such theory could be constructed, or that the general approach was useless. The realization that I came to is nicely summarized in an epitaph by Langer (1962: ix), quoted by W. Tecumseh Fitch in the same issue of *Biolinguistics* in which L&N's paper appeared:

> The chance that the key ideas of any professional scholar's work are pure non-sense is small; much greater the chance that a devastating refutation is based on a superficial reading or even a distorted one, subconsciously twisted by a desire to refute.   (Fitch 2009: 286)

Rather than dismissing the ideas of Kripke, Putnam, Dummett, Burge, Lewis, and other professional scholars as stupid or useless, I decided to attempt to explain their insights in a way that was consistent with my understanding of language, as a linguist trained in both theoretical (Chomskyan) and socio-historical linguistics. Lassiter (2008) was the result. Whatever the value of that specific proposal, I simply do not accept that this project is a waste of time in principle, as L&N seem to believe. Judging by his inclusion of the above quote, at least one prominent biolinguist would agree.

Fitch quotes the above passage in a section where he discusses sociological barriers to progress in biolinguistics, in particular citing "terminology and differing conceptions of 'language'" as impediments. As the discussion in section 2 suggests, I am of the opinion that L&N's critique of my work, like a great deal of the internalism–externalism debate more generally, deals more with these sociological issues than with genuine theoretical problems. The only real point of contention that I can identify is L&N's insistence (echoing Chomsky) that there can be no scientific theory of language use. I am unable to locate a substantive argument for this negative claim in L&N or in any of Chomsky's work; and, if we were to accept it, this would mean

dismissing existing efforts to provide just such a theory—within pragmatics and sociolinguistics, as well as related work in psychology, anthropology, economics, political science, and other disciplines—as "pure nonsense".

It is much better, I think, for internalists—really, I mean biolinguists, of whom I consider myself one—to try to integrate their theories with neighboring disciplines, and whenever possible to make use of other theories' insights in our own work. Here again Fitch's discussion of how to make progress in biolinguistics is relevant:

> Theoretical discussions are often dominated by rhetorical battles and ideological or terminological debate rather than constructive attempts to make tangible progress. Much of the criticism that currently divides the relevant fields boils down to "My opponent says we should look to *x* for answers, but I believe we should look to *y* instead". Typically, both *x* and *y* are probably important. Given the large number of open questions, biolinguistics will be better off when individual researchers pursue those topics and approaches they believe are important and promising, and refrain from attacking others who have different interests or try different approaches. There is little to be gained from such attacks, and if my experience is any guide, much to be lost.   (Fitch 2009: 291)

I think that both internalist and externalist aspects of language are important, and that both will eventually admit of scientific explanation (though probably not in the same terms). Anyone who wants to disagree with either my interest or my optimism should of course feel free; but it makes no sense to deny others the opportunity to look for answers, simply because one personally does not think they will be found. In any case, the proof is in the pudding: As we have seen, many scholars have already developed sophisticated approaches to language use. I suspect that useful and predictive theories of language use will continue to develop, as they have already been developing for many years, uninhibited by the insistence of certain internalists that this simply cannot happen.

What would the place of externalist inquiry be within biolinguistics? An important tradition in sociolinguistics, associated in particular with the work of Dell Hymes (e.g., Hymes 1974), deals with the notion of 'communicative competence'. Hymes suggests that the ability to use language in socially appropriate ways is part of individuals' knowledge of language, broadly construed. A great deal of empirical and experimental study in sociolinguistics has confirmed that speakers have detailed knowledge, not just of grammatical features of language, but also of *how to use language appropriately in a social context.* Biolinguists who make use of the model of Hauser *et al.* (2002) need not dismiss this notion as "unscientific"—rather, communicative competence can be treated as part of the 'faculty of language in the broad sense' (FLB), that is, as one of various areas where non-linguistic cognitive skills interface with grammar narrowly construed. Likewise, pragmatics both in the tradition of Grice (1989) and the more recent attempts to provide game-theoretic foundations for pragmatics (Parikh 2001, Benz *et al.* 2006) treats language use as the interface between purely linguistic knowledge and domain-general social reasoning and decision-making abilities. (The latter, by the way, also provides a precise, testable,

and mathematically rigorous theory of language use, again contrary to the claim that no such theory is possible.)

I see no conflict between externalist lines of inquiry like these and internalism, beyond the exclusionary attitude of some internalists against those who pursue other approaches to language. Rather, they fit naturally within the framework proposed by Hauser *et al*. (2002) as aspects of FLB. My attempt in Lassiter (2008) to articulate a notion of 'social meaning' was an extension of these aspects of FLB: In particular, I focused on word meaning as it related both to grammar (the narrow faculty) and pragmatic and communicative competence as social skills. Communicative competence is an inherently social skill, and so it comes as no surprise that, if two language users with identical internal states were embedded in different social environments, one might be judged communicatively competent while the other is not. This is really all that Burge's 'arthritis' story shows: Our intuitions about what constitutes appropriate language use does not depend exclusively on the internal states of language users, but also on who they are in communication with and whether their present internal states will allow for successful communication. Many philosophers, in particular, tend to include such facts in their notion of 'meaning', a practice which I followed in Lassiter (2008). This may not agree with some linguists' use, but it is a mere terminological issue, and perfectly acceptable usage if flagged appropriately.

The notion of social meaning is simply not a challenge to internalist explanations of linguistic competence. The situation is the same as it would be, say, in decision theory: The claim that individual agents attempt to make the best possible choices given their information and preferences is not threatened by the fact that agents sometimes make mistakes due to bad information or processing constraints (see Gintis 2009 for discussion of the competence–performance distinction in this domain). The distinction between the mechanism and how the mechanism is put to use in concrete circumstances is not threatened by the existence of research attempting to connect the two. Further, my construal of notions like 'normativity of meaning', 'deference', etc. was an attempt to embed a theory of individual agents' internal states within a social context and derive predictions. L&N and anyone else should, of course, feel free to disagree with or ignore this attempt; but it is fundamentally opposed to the spirit of scientific inquiry—not to mention bad for the future of the science to which this journal is dedicated to cultivating—for L&N to trivialize the effort or discourage further investigation simply because it is not what *they* do, or because they personally dislike such efforts.

More generally, linguists simply cannot afford to ignore issues of language use. Those who do will be deprived of the insights of highly relevant and increasingly intricate theories—from the realms of pragmatics, psychology, and computational modeling, to name a few—many of which are compatible with, or even presuppose, the biolinguistic thesis that language acquisition and structure are explained in large part by biological facts about humans. If biolinguists ignore or trivialize these efforts on the basis of *a priori* claims about what kinds of theories are possible, I fear that biolinguistics itself will come to be ignored and trivialized as

theories of the supposedly impossible type are actually developed. It would be better to give up this provincialism and acknowledge that there is no principled conflict between biolinguistics and theorizing about language use—in other words, that there is no irreconcilable conflict between internalism and externalism.

References

Benz, Anton, Gerhard Jäger & Robert van Rooij (eds.). 2006. *Game Theory and Pragmatics.* Basingstoke: Palgrave Macmillan.

Bicchieri, Cristina, Richard Jeffrey, and Brian Skyrms. 1996. *The Dynamics of Norms.* Cambridge: Cambridge University Press.

Bicchieri, Cristina. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge: Cambridge University Press.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use.* New York: Praeger.

Chomsky, Noam. 1995. Language and nature. *Mind* 104, 1–61.

Chomsky, Noam. *New Horizons in the Study of Language and Mind.* Cambridge: Cambridge University Press.

Clark, Herbert H. 1992. *Arenas of Language Use.* Chicago: University of Chicago Press.

Clark, Herbert H. 1996. *Using Language.* Cambridge: Cambridge University Press.

Dawkins, Richard. 1986. *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design.* New York: Norton.

Eckert, Penelope & John R. Rickford. 2001. *Style and Sociolinguistic Variation.* Cambridge: Cambridge University Press.

Epstein, Joshua & Robert Axtell. 1996. *Growing Artificial Societies: Social Science from the Bottom Up.* Washington, DC: Brookings Institution Press & Cambridge, MA: MIT Press.

Epstein, Joshua. 2007. *Generative Social Science: Studies in Agent-Based Computational Modeling.* Princeton, NJ: Princeton University Press.

Fitch, W. Tecumseh. 2009. Prolegomena to a future science of biolinguistics. *Biolinguistics* 3, 321–331.

Giles, Howard & Peter F. Powesland. 1975. *Speech Style and Social Evaluation.* London: Academic Press.

Giles, Howard & W. Peter Robinson (eds.). 1990. *Handbook of Language and Social Psychology.* New York: Wiley.

Gintis, Herbert. 2009. *The Bounds of Reason.* Princeton, NJ: Princeton University Press.

Grice, Herbert Paul. 1989. *Studies in the Ways of Words.* Cambridge, MA: Harvard University Press.

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

Hinzen, Wolfram. 2006. Internalism about truth. *Mind and Society* 5, 139–166.

Horn, Laurence. 1989. *A Natural History of Negation.* Chicago: University of Chicago

Press.

Hymes, Dell. 1974. *Foundations in Sociolinguistics: An Ethnographic Approach.* Philadelphia, PA: University of Pennsylvania Press.

Labov, William. 1994. *Internal Factors — Principles of Linguistic Change*, vol. I. Malden, MA: Blackwell.

Labov, William. 2001. *Social Factors — Principles of Linguistic Change*, vol. II. Malden, MA: Blackwell.

Lassiter, Daniel. 2008. Semantic externalism, language variation, and sociolinguistic accommodation. *Mind and Language* 23, 607–633.

Lewis, David. 1969. *Convention.* Cambridge, MA: Harvard University Press.

Lohndal, Terje & Hiroki Narita. 2009. Internalism as methodology. *Biolinguistics* 3, 321–331.

McGilvray, James. 2002. MOPs: The science of concepts. In Wolfram Hinzen & Hans Rott (eds.), *Belief and Meaning: Essays at the Interface*, 73–103. Frankfurt am Main: Ontos.

Narita, Hiroki. Forthcoming. The naturalist program for neo-Cartesian biolinguistics. *Proceedings of the 4th Annual Meeting of the Sophia University Linguistic Society.*

Le Page, Robert B., and André Tabouret-Keller. 1985. *Acts of Identity: Creole-Based Approaches to Language and Ethnicity.* Cambridge: Cambridge University Press.

Parikh, Prashant. 2001. *The Use of Language.* Stanford, CA: CSLI Publications.

Payne, Arvilla. 1980. Factors controlling the acquisition of the Philadelphia dialect by out-of-state children. In William Labov (ed.), *Locating Language in Time and Space*, 143–178. New York: Academic Press.

van Rooij, Robert. 2003. Questioning to resolve decision problems. *Linguistics and Philosophy* 26, 727–763.

Skyrms, Brian. 1996. *The Evolution of the Social Contract.* Cambridge: Cambridge University Press.

Skyrms, Brian. 2004. *The Stag Hunt and the Evolution of Social Structure.* Cambridge: Cambridge University Press.

Sperber, Dan & Deirdre Wilson. 1986. *Relevance: Communication and Cognition.* Cambridge, MA: Harvard University Press.

Young, H. Peyton. 1998. *Individual Strategy and Social Structure.* Princeton, NJ: Princeton University Press.

Daniel Lassiter
New York University
Department of Linguistics
*10 Washington Place*
New York, NY 10003
USA
lassiter@nyu.edu

The Biolinguistics Network

Anna Maria Di Sciullo

This note highlights the Biolinguistics Network, its creation, and its role as a promising avenue of research in the biological basis of the language faculty. It also provides insights into the type of material discussed by the Network's participants so far, and the questions that will be addressed in upcoming events. The Biolinguistics Network was created in 2007 to foster multidisciplinary research by setting up a dynamic space to address biolinguistic questions, including what are the principles of our knowledge of language, how this knowledge grows, how it is put to use, how it evolved, and which aspects of the machinery are unique to language as opposed to shared with other domains of knowledge. Such questions were discussed in the two conferences that led to the creation of the Biolinguistics Network. The first, *Biolinguistic Investigations*, took place in Santo Domingo in February, 2007. The second, *Biolinguistic Perspectives on Language Evolution and Variation*, was held in Venice in June of that same year. These meetings brought together a number of contributors to the field. Selected papers from these two conferences are assembled in *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty* (Di Sciullo & Boeckx, in press). Such events and related publications, including the multi-authored cross-disciplinary piece published in this issue (Di Sciullo *et al.* 2010), are exemplars of the catalytic role of the Biolinguistics Network in our understanding of the biology of language.

I

The first meeting of the network was held February 23–24, 2008, at the University of Arizona at Tucson. This meeting was organized jointly with Massimo Piattelli-Palmarini and sponsored by the Social Sciences and Humanities Research Council of Canada, the Université du Québec à Montréal, and the University of Arizona at Tucson.
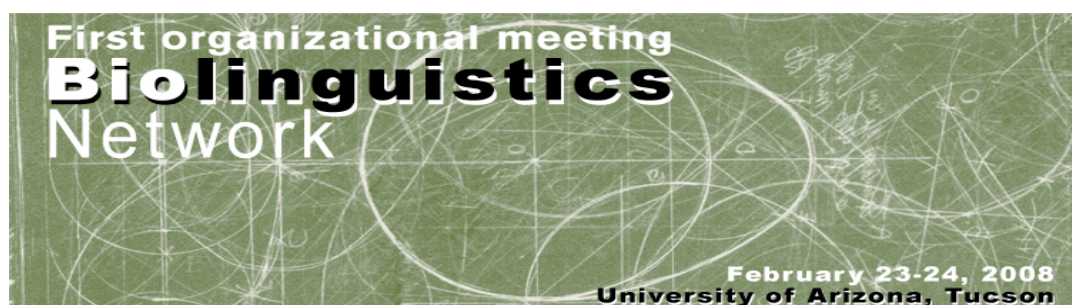
This meeting set out the architecture of the Biolinguistics Network, including possible websites, journals, summer schools, permanent centers, and primary roles. The network website www.biolinguistics.uqam.ca was created shortly after the Arizona meeting providing the names and web pages of the participants of the network, the links to preceding and forthcoming Biolinguistic conferences, and a thematic bibliography with downloadable papers.

The Arizona meeting provided a great venue for the presentation of current work and research programs in the understanding of the faculty of language (FLB and FLN, in the sense of Hauser *et al.* 2002) and its biological basis. The material presented at this meeting was incredibly rich and covered a wide range of biolinguistic questions. I will summarize these contributions here, referring the reader to the Biolinguistics Network website for the abstracts of each presentation.

FLB

The first session of the meeting addressed basic questions on the properties of the faculty of language in the broad sense (FLB). Marc Hauser and Ansgar Endress discussed the constraints imposed by general mechanisms of perception that influence the form that grammars can take, and reported experimental results on temporal ordering rules and edge constraints, with parallel results in humans and chimps. Tom Bever discussed some aspects of the relations between biology, learning, and evolutionary constraints on linguistic universals. He presented results on sequence learning, language learning/behavior, and fMRI studies of language behavior. Karin Stromswold talked about the 'critical period' in language acquisition and addressed three questions on the genetic, cognitive and neural bases of the 'critical period' in language acquisition. The first question is: What is the critical period? The second is: What causes the critical period? And the third question is: Why is there a critical period? Ken Wexler discussed optional infinitives in language acquisition, as well as behavioral genetic research (twin studies) showing that, even in typical development (having to do with syntactic control), there appears to be genetically linked variation. The general point of Wexler's presentation was that genetics might bear on the development and inheritance of particular properties of language, not just on language as a whole.

Genes

The second section of the conference focused on genes, and in particular on our understanding of the role of the *FOXP2* gene, previously thought by some as being the 'language gene'. Lyle Jenkins pointed out that interdisciplinary work in a large number of subfields of biolinguistics in recent years has provided a preliminary and partial sketch of the mapping of the *FOXP2* gene to its phenotype (including language and cognition). However, he noted that the following questions remain unanswered. What exactly is the phenotype of the developmental verbal dyspraxia caused by *FOXP2*? How can we sharpen the distinction between the linguistic features of the phenotype from other cognitive problems? How can we determine what genes involved in language are downstream from *FOXP2*? And what can we learn about the brain areas and circuits sub-served by

language? Robert Berwick suggested three separate lines of experiments regarding *FOXP2* in order to extend the domain of evidence, relating *FOXP2* to sign language and motor serialization, to evolution, and to the 'songeme' analysis of normal and knock-down zebra finches. Massimo Piattelli-Palmarini and Juan Uriagereka's talk, 'Spellbound: The birds', suggested that the narrow language faculty may have evolved as a procedure that 'externalizes' a more ancient form of recursion. To address the parsing problem that results, they discussed the role of songbirds' Area X, the functional equivalent of our caudate nucleus, which is responsible for the acquisition and the production of the singing behavior, and the role of songbird *FoxP2* in this area. In his second talk, Robert Berwick also pointed out that even though recent research has suggested that *FOXP2* has undergone strong natural selection within the last 100,000 years, the differences we observe between human *FOXP2* and *FoxP2* in non-human primates could just as easily be due to chance alone.

FLN to FLB

The last section of the conference was devoted to the bridging FLN to FLB. I focused on linking the basic Minimalist operator, Merge, to experimental results. I presented the results of behavioral and ERP experiments indicating that there are differences in the processing of morphological expressions with the same linear properties but different in their hierarchical structure. I proposed that "deep asymmetry effects" follow from the computation of Merge by the brain. Cedric Boeckx stressed the need for biolinguists to take very seriously the recent (minimalist) trend that seeks to "approach UG from below" and uncover basic, primitive operations, such as Merge. He focused on the syllable and the phase, and argued that the minimalist search for basic relations is the best bet we have for formulating 'linking hypotheses' in the domain of neurolinguistics and evolutionary linguistics. Lisa Cheng's presentation focused on syntax–phonology (mis)matches and their mappings at different interfaces. The questions she considered included the following: How is syntactic phrasing mapped with phonological phrasing? Or how does syntactic phrasing feed phonological phrasing? One extension of this program is at the domain of psycho-/neuro-linguistics: How does sentence planning work, taking into consideration the syntax–phonology interface? Elly van Gelderen discussed which insights historical syntax provides regarding economy principles, when couched within a biolinguistic framework. She raised the following question: How is cyclical change from analytic to synthetic to analytic relevant to the claim that "the conflict between computational efficiency and ease of communication" is resolved "to satisfy the CI interface"? Heidi Harley & Andrew Wedel presented work on the mapping of the hierarchical structure of syntax onto the linear structure of the utterance. They investigated this mapping by contrasting the cross-linguistically uniform selectional/hierarchical relationship between verbs and their objects with the cross-linguistically varying temporal relationship between verbs and their objects in SOV and SVO languages. Jim McGilvray discussed linguistically expressed concepts (HOUSE, TIGER, MOST, SILLY, WASH…), pointing out that the study of such concepts requires the kind of internal machinery that exploits resources

within the mind/head, each of which develops according to some internally set agenda. This agenda is multifaceted, and calls upon the study of internal systems (computational theories of vision, for example), developmental data, evidence concerning how linguistic concepts are employed by/within other systems, and evo–devo materials.

In addition to the setting of the organizational aspect of the Biolinguistics Network, two publications came out of the discussions of the first meeting of the network: Hauser & Bever (2008) and Di Sciullo *et al.* (2010). These papers resulted from a collective effort to address biolinguistic questions, to showcase the advances, and to point to open questions for a larger public, encompassing linguistics and biology, as well as cognitive and computational sciences.

II

The upcoming meeting of the Biolinguistics Network, *The Language Design* conference, will be held this spring at the Université du Québec à Montréal (May 27–30, 2010). Its purpose is to address further biolinguistic questions, and in particular the factors entering into the human language design stemming from linguistic theory, biolinguistics, and biophysics. This workshop will bring together participants from a broad array of disciplines to discuss topics that include the connection between linguistic theory and genetics, evolutionary developmental biology and language variation, and computer science/information theory and the reduction of uncertainty/complexity. While the first meeting of the network addressed general questions on FLB, FLN–FLB, and genetics, the questions raised by the forthcoming event are more fine-grained. The themes that will be discussed include the role of properties of grammars, operations, relations, complexity reducing factors, brain studies, as well as studies in language variation and change on our understanding of the biology of language. The following paragraphs outline some of the questions that will be addressed.

Formal Grammars and Human/Animal Comparative Studies

Recent work on Chomsky's hierarchy of formal grammars, as well as on formal grammars and human/animal comparative studies, brings to the fore issues that go back to the 1950s on the sort of grammars/automata that specifically describe/generate human language (Chomsky 1956). Several questions arise, including whether more than one sort of formal grammar is part of the language design, perhaps distributed in different components (Bergelson & Idsardi 2009), or possibly available within narrow syntax itself (Lasnik, in press). Recent human/animal comparative studies on learnability raise similar questions (Fitch & Hauser 2002, Jarvis 2004, 2006, Saar & Mitra 2008, Friederici 2009). To what extent do such studies shed light on the specificity of human language?

The Properties of the Operators of FL(N)

One important question is the nature of the operators that derive the discrete infinity of human language. Merge and Agree are assumed to be the dyadic operators of the Faculty of Language. However, several questions arise with

respect to their properties. Is Merge completely free (Boeckx, to appear), or is it subject to formal conditions (Frampton & Gutmann 2002)? Further questions arise on the nature of the representations derived by Merge, and whether or not these are restricted to adjunction (Hornstein & Pietroski, forthcoming). More questions arise regarding the semantics of human language, the kinds of operators that derive the interpretation, and whether the derivations are external to narrow syntax (Pietroski 2008, Higginbotham 2009, Hinzen, in press, to appear). What kind of biological evidence would support the theoretical in this regard?

The Role of (A)Symmetry in Grammar and Biology

Properties of relations such as symmetry, asymmetry and antisymmetry have been shown to be relevant in the language design. Symmetry-breaking has been proposed to drive derivations (Moro 2000, 2008) and to account for word-order differences (SVO, VSO, etc.) (Jenkins 2000, 2004, in press a, in press b); antisymmetry has been argued to be a central property in syntax, as well as for linearization (Kayne 1994), and asymmetry has been claimed to be part of Merge (Chomsky 1995, Di Sciullo 2003a, 2003b, 2005, Zwart 2006, Di Sciullo & Isac 2008a, 2008b). We know that properties of relations are used to describe the dynamics of morphogenesis in biology (Montell 2008), and to formulate laws of physics. Why should these abstract properties of form participate in the language design? What is the basis of their dynamics in human language?

Complexity-Reducing Factors in Derivations, Biology, and Physics

It is generally assumed, since Chomsky's (2005) three factors, that the factors reducing derivational complexity are external to the language design. They include mechanisms that reduce the search space and the choice points in the derivations. Phases are part of the factors reducing derivational complexity in narrow syntax (Uriagereka 1999, Chomsky 2001, 2008, Boeckx & Grohmann 2007, Grohmann 2009). Other complexity-reducing factors include the mechanisms restricting the set of possible acquirable grammars (Yang 2002, Di Sciullo & Fong 2005, Niyogi 2006, Roeper 2007), those that reduce the set of possible interpretations for linguistic representations (Reinhart 2006, Speas, in press), and those that come from limits imposed by perception and memory (Chomsky & Miller 1963, Bever 1970). Are these computational constraints related to one another? Are there correlates to complexity-reducing factors in biology or in physics?

Variation in Languages and in Biology

Advances in our understanding of language variation since Principles–and–Parameters have made it possible to derive observable differences between languages from abstract properties of the grammar and phylogenetics (Guardiano & Longobardi 2005, Longobardi & Guardiano, in press). Recent findings in the dynamics of morphogenesis, regulatory HOX genes (Gehring & Ikeo 1999, Gehring 2005), and philogenetic patterns of variance (Palmer 2004a, 2004b, 2009) are interesting from a biolinguistic perspective (Niyogi & Berwick 1997, 2009, Di Sciullo, in press). They point at the central role of asymmetry in the dynamics of variation and change in the biological world. How can our knowledge of vari-

ation and change in genetics and population biology enhance our understanding of language diversity?

Genetics, Brain Studies, and Language Impairments

From an evolutionary and comparative standpoint, *FOXP2* has been intensely analyzed as potentially shedding light on the unique characteristics of the human species, as well as on human origins. However, given *FOXP2*'s multifactorial neural influence and its role as part of the externalization system for language, it would then seem speculative at this point to base strong conclusions on such evidence (Berwick & Chomsky, in press and Piattelli-Palmarini & Uriagereka, in press). How do advances in our understanding of SLI and other genetically endowed language impairments, such as Williams–Beuren syndrome (Perovic & Wexler 2007), as well as advances in the study of brain-level mechanisms that support language (Fiorentino & Poeppel 2007, Lau *et al.* 2008), shed light on the language design?

The conference will offer a renewed opportunity to bring forward new approaches to these important questions. The confirmed speakers are: Robert Berwick, Cedric Boeckx, Roberto De Almeida, Anna Maria Di Sciullo, Sandiway Fong, Jason Ginsberg, Kleanthes Grohmann, Wolfram Hinzen, James Higginbotham, William Idsardi, Dana Isac, Lyle Jenkins, Howard Lasnik, Giuseppe Longobardi, Partha Mitra, Richard Palmer, Massimo Piattelli-Palmarini, Paul Pietroski, David Poeppel, Charles Reiss, Tom Roeper, Margaret Speas, Juan Uriagereka, and Kenneth Wexler. The abstracts will be posted shortly on the network website.

III

The Biolinguistics Network keeps track of the participants, the publications and the conferences in this field. Together with the *Biolinguistics* journal and the recently created *Biolinguistics* blog, it strengthens the links between individuals, groups and institutions interested in pushing forward our knowledge of the biology of language.

As the founder of the Biolinguistics Network, I believe that in such a dynamic space, it is likely that the biolinguistic questions will meet explanatory hypotheses and that new insights will lead to a better understanding of language as a natural object.

References

Bergelson, Elika & William J. Idsardi. 2009. Structural biases in phonology: Infant and adult evidence from artificial language learning. *BUCLD* 33, 85–96.

Berwick, Robert C. & Noam Chomsky. In press. The biolinguistic program: The current state of its evolution and development. In Di Sciullo & Boeckx (eds.).

Berwick, Robert C. In press. *Syntax facit saltum redux*: Biolinguistics and the leap

to syntax. In Di Sciullo & Boeckx (eds.).

Bever, Thomas G. 1970. The cognitive basis for linguistic structures. In John R. Hayes (ed.), *Cognition & the Development of Language*, 279–362. New York: John Wiley.

Boeckx, Cedric. To appear. A tale of two minimalisms — Reflections on the plausibility of crash-proof syntax and its free merge alternative. In Michael Putnam (ed.), *Exploring Crash-Proof Grammars*. Amsterdam: John Benjamins.

Boeckx, Cedric & Kleanthes K. Grohmann. 2007. Remark: Putting phases in perspective. *Syntax* 10, 294–222.

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2, 113–23.

Chomsky, Noam. 1995. *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Robert Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik*, 89–155. Cambridge, MA: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In Michael Kenstowicz (ed.), *Ken Hale: A Life in Language,* 1–52. Cambridge, MA: MIT Press.

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

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

Chomsky, Noam & Charles Miller. 1963. Finitary models of language users. In R. Duncan Luce, Robert R. Bush & Eugene Galanter (eds.), *Handbook of Mathematical Psychology*, vol. II, 419–491. New York: Wiley.

Di Sciullo, Anna Maria (ed.). 2003a. *Asymmetry in Grammar*, vol. 1 — *Syntax and Semantics*. Amsterdam: John Benjamins.

Di Sciullo, Anna Maria (ed.). 2003b. *Asymmetry in Grammar*, vol. 2 — *Phonology, Morphology, and Acquisition*. Amsterdam: John Benjamins.

Di Sciullo, Anna Maria. 2005. *Asymmetry in Morphology*. Cambridge, MA: MIT Press.

Di Sciullo, Anna Maria. In press. A biolinguistic approach to variation. In Di Sciullo & Boeckx (eds.).

Di Sciullo, Anna Maria & Cedric Boeckx (eds.). In press. *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*. Oxford: Oxford University Press.

Di Sciullo, Anna Maria & Sandiway Fong. 2005. Morpho-syntax parsing. In Anna Maria Di Sciullo (ed.), *UG and External Systems*, 247–268. Amsterdam: John Benjamins.

Di Sciullo, Anna Maria & Dana Isac. 2008a. The asymmetry of merge. *Biolinguistics* 2, 260–290.

Di Sciullo, Anna Maria & Dana Isac. 2008b. Movement chains at the interfaces. *The Canadian Journal of Linguistics/Revue Canadienne de Linguistique* 53, 181–219.

Di Sciullo, Anna Maria, Massimo Piattelli-Palmarini, Kenneth Wexler, Robert C. Berwick, Cedric Boeckx, Lyle Jenkins, Juan Uriagereka, Karin Stromswold, Lisa Lai-Shen Cheng, Heidi Harley, Andrew Wedel, James McGilvray, Elly

van Gelderen & Tom Bever. 2010. The biological nature of language. *Biolinguistics* 4, 4–34.

Fiorentino, Robert & David Poeppel. 2007. Compound words and structure in the lexicon. *Language and Cognitive Processes* 22, 953–1000.

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

Frampton, John & Sam Gutmann. 2002. Crash-proof syntax. In Samuel David Epstein & T. Daniel Seely (eds.), *Derivation and Explanation in the Minimalist Program*, 90–105. Malden, MA: Blackwell.

Friederici, Angela D. 2009. The brain differentiates hierarchical and probabilistic grammars. In Piattelli-Palmarini, Uriagereka & Salaburu (eds.), 184–194.

Gehring, Walter J. 2005. New perspectives on eye development and the evolution of eyes and photoreceptors. *Journal of Heredity* 96, 171–184.

Gehring, Walter J. & Kazuko Ikeo. 1999. Pax 6: Mastering eye morphogenesis and eye evolution. *Trends in Genetics* 15, 371–377.

Grohmann, Kleanthes K. (ed.). 2009. *InterPhases: Phase-Theoretical Investigations of Linguistic Interfaces*. Oxford: Oxford University Press.

Guardiano Cristina & Giuseppe Longobardi. 2005. Parametric comparison and language taxonomy. In Montserat Batllori, Maria-Luïsa Hernanz, Carme Picallo & Francesc Roca (eds.), *Grammaticalization and Parametric Variation*, 149–174. Oxford: Oxford University Press.

Hauser, Marc D. & Thomas G. Bever. 2008. A biolinguistic agenda. *Science* 322, 1057–1059.

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

Higginbotham, James. 2009. Two interfaces. In Piatelli-Palmarini, Uriagereka & Salaburu (eds.), 142–150.

Hinzen, Wolfram. In press. Emergence of a systemic semantics through minimal and underspecified codes. In Di Sciullo & Boeckx (eds.).

Hinzen, Wolfram. To appear. Minimalism. In Ruth Kempson, Nicholas Asher & Tim Fernando (eds.), *Handbook of the Philosophy of Linguistics*. Amsterdam: Elsevier.

Hornstein, Norbert & Paul Pietroski. Forthcoming. Basic operations: Minimal syntax–semantics. *Catalan Journal of Linguistics.*

Jarvis, Erich D. 2004. Learned birdsong and the neurobiology of human language. *Annals of the New York Academy of Science* 1016, 749–777.

Jarvis, Erich D. 2006. Selection for or against vocal learning in birds and mammals. *Ornithological Science* 5, 5–14.

Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press.

Jenkins, Lyle (ed.). 2004. *Variation and Universals in Biolinguistics*. Amsterdam: Elsevier.

Jenkins, Lyle. In press a. Biolinguistic investigations: Genetics and dynamics. In Di Sciullo & Boeckx (eds.).

Jenkins, Lyle. In press b. Biolinguistics: The 'third factor' in evolution and variation. In Di Sciullo & Boeckx (eds.).

Kayne, Richard S. 1994. *The Antisymmetry of Syntax*. Cambridge, MA: MIT Press.

Lasnik, Howard. In press. What kind of computing device is the human language faculty? In Di Sciullo & Boeckx (eds.).

Lau, Elen, Colin Phillips & David Poeppel. 2008. A cortical network for semantics: (De)constructing the N400. *Nature Reviews Neuroscience* 9, 920–933.

Longobardi, Giuseppe & Cristina Guardiano. In press. The biolinguistic program and historical reconstruction. In Di Sciullo & Boeckx (eds.).

Montell, Denise J. 2008. Morphogenetic cell movements: Diversity from modular mechanical properties. *Science* 322, 1502–1505.

Moro, Andrea. 2000. *Dynamic Antisymmetry*. Cambridge, MA: MIT Press.

Moro, Andrea. 2008. *The Boundaries of Babel: The Brain and the Enigma of Impossible Languages*. Cambridge, MA: MIT Press.

Niyogi, Partha. 2006. *The Computational Nature of Learning and Evolution*. Cambridge, MA: MIT Press.

Niyogi Partha & Robert C. Berwick. 1997. A dynamical systems model of language change. *Complex Systems* 11, 161–204.

Niyogi, Partha & Robert C. Berwick. 2009. The proper treatment of language acquisition and change in a population setting. *Proceedings of the National Academy of Sciences of the United States of America* 106, 10124–10129.

Palmer, A. Richard. 2004a. Symmetry breaking and the evolution of development. *Science* 306, 828–833.

Palmer, A. Richard. 2004b. Antisymmetry. In Benedikt Hallgrimson & Brian K. Hall (eds.), *Variation: A Central Concept in Biology*, 359–390. Amsterdam: Elsevier.

Palmer, A. Richard. 2009. Animal asymmetry. *Current Biology* 19, R473.

Perovic, Alexandra & Kenneth Wexler. 2007. Complex grammar in Williams syndrome. *Clinical Linguistics & Phonetics* 22, 729–745.

Piattelli-Palmarini, Massimo & Juan Uriagereka. In press. A geneticist's dream, a linguist's nightmare: The case of FoxP2. In Di Sciullo & Boeckx (eds.).

Piattelli-Palmarini, Massimo, Juan Uriagereka & Pello Salaburu (eds.). 2009. *Of Minds and Language: A Dialogue with Noam Chomsky in the Basque Country*. Oxford: Oxford University Press.

Pietroski, Paul. 2008. Minimalist meaning, internalist interpretation. *Biolinguistics* 4, 317–41.

Reinhart, Tanya. 2006. *Interface Strategies*. Cambridge, MA: MIT Press.

Roeper, Tom. 2007. *The Prism of Grammar*. Cambridge, MA: MIT Press.

Saar, Sigal & Partha Mitra. 2008. A technique for characterizing the development of rhythms in bird song. *Public Library of Science One* 3, e1461.

Speas, Margaret. In press. Generalized functional heads. In Anna Maria Di Sciullo & Virginia Hill (eds.), *Edges, Heads and Projections: Interface Properties*. Amsterdam: John Benjamins.

Stromswold, Karin. 2008. The genetics of speech and language impairments. *The New England Journal of Medicine* 359, 2381–2383.

Stromswold, Karin. 2010. Genetics and the evolution of language: What genetic studies reveal about the evolution of language. In Richard K. Larson, Viviane Déprez & Hiroko Yamakido (eds.), *Evolution of Language: Biolinguistic Perspectives*, 176–193. Cambridge: Cambridge University Press.

Uriagereka, Juan. 1999. Multiple spell-out. In Samuel David Epstein & Norbert Hornstein (eds.), *Working Minimalism*, 251–283. Cambridge, MA: MIT Press.

Wexler, Kenneth. 1996. The development of inflection in a biologically based theory of language acquisition. In Mabel L. Rice (ed.), *Toward a Genetics of Language*, 113–144. Mahwah, NJ: Lawrence Erlbaum.

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

Yang, Charles. 2002. *Knowledge and Learning in Natural Language*. Oxford: Oxford University Press.

Yang, Charles. 2005. The origin of linguistic irregularity. In James W. Minett & William S.-Y. Wang (eds.), *Language Acquisition, Change and Emergence: Essays in Evolutionary Linguistics*, 296–328. Hong Kong: City University of Hong Kong Press.

Yang, Charles. In press. Three factors in language variation. In Di Sciullo & Boeckx (eds.).

Zwart, C. Jan-Wouter. 2006. Local agreement. In Cedric Boeckx (ed.), *Agreement Systems*, 317–339. Amsterdam: John Benjamins.

Anna Maria Di Sciullo
Université du Québec à Montréal
Department of linguistics
P.O. Box 8888, Downtown Station
Montréal, QC H3C 3P8
Canada
di_sciullo.anne-marie@uqam.ca

TABLE OF CONTENTS

★ EDITORIAL ★

★ BRIEFS ★

★ REVIEWS ★

★ FORUM ★