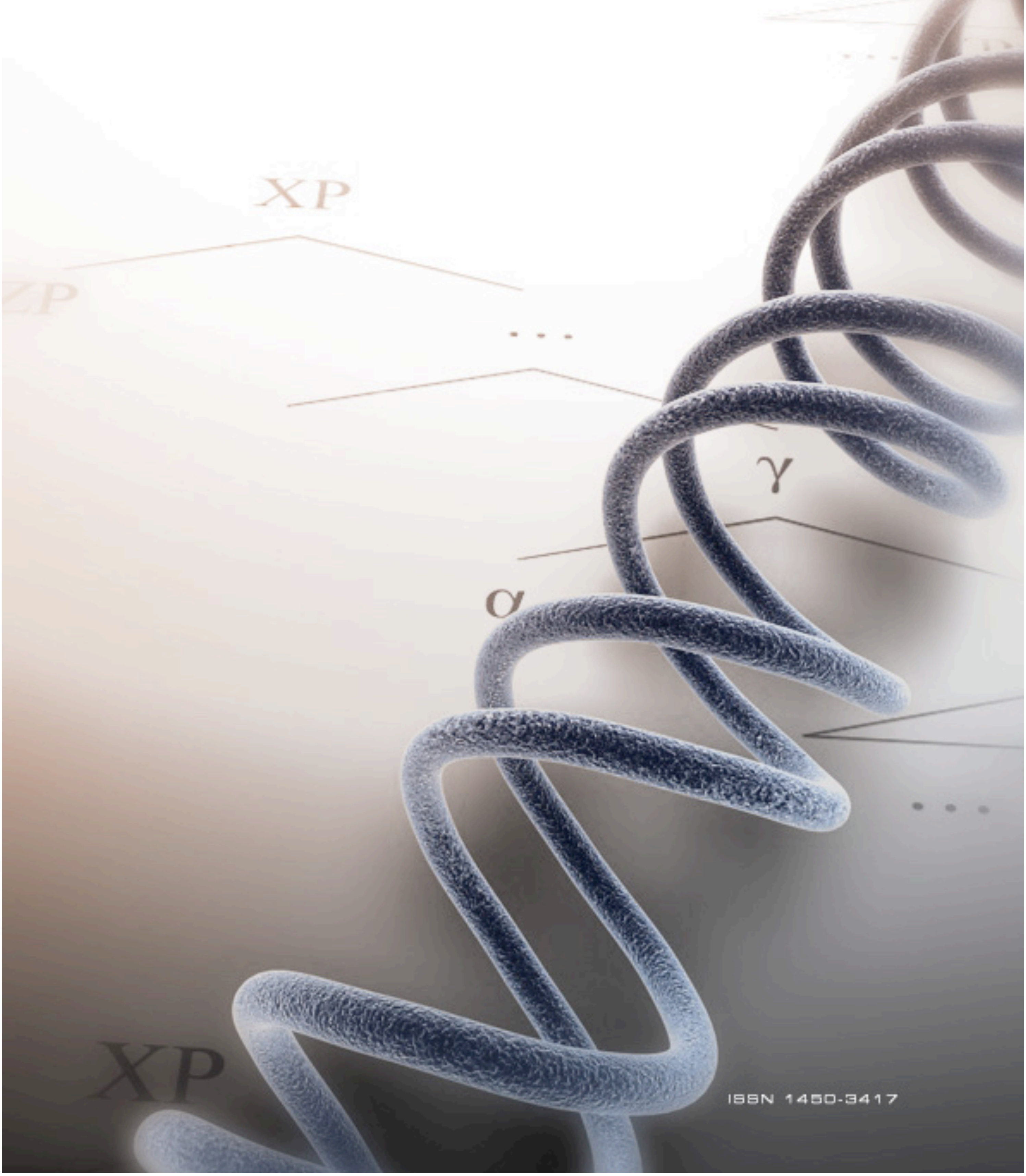
biolinguistics

volume 1 fall 2007



ISSN 1450-3417

★ EDITORIAL ★

The *Biolinguistics* Manifesto

Cedric Boeckx & Kleanthes K. Grohmann

1. What We Mean by *Biolinguistics*

Exactly fifty years ago Noam Chomsky published *Syntactic Structures* (Chomsky 1957), a slim volume that conveyed some essential results of his then unpublished *Logical Structure of Linguistic Theory* (Chomsky 1955/1975). The results were presented in such a way as to emphasize key aspects of the combinatorial properties of grammar (a reflex of the fact that the volume grew out of class notes for an audience of engineers), but, as is well-known, *Syntactic Structures* had an important subliminal message that was made explicit in Chomsky's famous review of Skinner's *Verbal Behavior* (Chomsky 1959), and even more so in chapter 1 of *Aspects of the Theory of Syntax* (Chomsky 1965). The message, decidedly psychological in character, defines the central goal of the generative enterprise as that of identifying the properties of the human language faculty. This central goal can be broken down into a series of more precise questions (see Chomsky 1986, 1988):

1. What is knowledge of language?
2. How is that knowledge acquired?
3. How is that knowledge put to use?
4. How is that knowledge implemented in the brain?
5. How did that knowledge emerge in the species?

Today these five questions constitute the conceptual core and focus of inquiry in fields like theoretical linguistics (the traditional areas of syntax, semantics, morphology, phonology), pragmatics, first and second language acquisition, psycholinguistics, neurolinguistics, and beyond.

What these research questions emphasize is the fact that language can, and should, be studied like any other attribute of our species, and more specifically, as an organ of the mind/brain.

The past fifty years have shown, uncontroversially in our opinion, that it makes eminent sense, at various levels, to regard the study of the language faculty as a branch of biology, at a suitable level of abstraction. After all, the five questions listed above are but (conceptually unpacked) variants of Tinbergen's famous four questions in his classic paper "On the Aims and Methods of Ethology" (Tinbergen 1963), a central document in the biology of (animal) behavior:

ISSN 1450–3417
Biolinguistics 1: 001–008, 2007
http://www.biolinguistics.eu

1. What stimulates the animal to respond with the behavior it displays, and what are the response mechanisms?
2. How does an organism develop as the individual matures?
3. Why is the behavior necessary for the animal's success and how does evolution act on that behavior?
4. How has a particular behavior evolved through time? Can we trace a common behavior of two species back to their common ancestor?

The goal of this new journal is to provide a forum, a context, and a framework for discussion of these foundational issues. We decided to call the journal *Biolinguistics* to highlight the commitment of the generative enterprise to the biological foundations of language, and to emphasize the necessarily interdisciplinary character of such enterprise.

There is both a weak and a strong sense to the term 'biolinguistics'. The weak sense of the term refers to "business as usual" for linguists, so to speak, to the extent they are seriously engaged in discovering the properties of grammar, in effect carrying out the research program Chomsky initiated in *Syntactic Structures*.

The strong sense of the term 'biolinguistics' refers to attempts to provide explicit answers to questions that necessarily require the combination of linguistic insights and insights from related disciplines (evolutionary biology, genetics, neurology, psychology, etc.). We regard Eric Lenneberg's book, *Biological Foundations of Language*, published exactly forty years ago (Lenneberg 1967), as the best example of research in biolinguistics in this strong sense.

We would like our journal to provide a forum for work in biolinguistics in both the weak and the strong sense. We would like to stress that the term 'weak sense' is not meant to indicate that we regard work focusing narrowly on properties of the grammar as inferior to interdisciplinary work. Indeed we think that such work is not only necessary, but has very often proven to be the basis for more interdisciplinary studies.

2. Why Start *Biolinguistics* Now?

The term 'biolinguistics' first appears, to our knowledge, as part of a book title, the *Handbook of Biolinguistics*, published nearly 60 years ago (Meader & Muyskens 1950). The book advocates (as the authors put it) a modern science of biolinguistics, whose practitioners "look upon language study […] as a natural science, and hence regard language as an integrated group of biological processes […]. This group seeks an explanation of all language phenomena in the functional integration of tissue and environment" (Meader & Muyskens 1950: 9).

The term 'biolinguistics' resurfaces in 1974 as part of a report on an interdisciplinary meeting on language and biology (Piattelli-Palmarini 1974), attended by Salvador Luria and Noam Chomsky, and organized by Massimo Piattelli-Palmarini, under the sponsorship of the Royaumont center for a Science of Man.

Around the same time (a period well-documented in Jenkins 2000), Lyle

Jenkins attempted to launch a journal entitled *Biolinguistics*, and received support from pre-eminent biologists (support documented by three extant letters reproduced in an Appendix to this editorial manifesto). The journal never materialized, but the concerns and issues discussed three decades ago didn't disappear. As a matter of fact, all these issues, many of which anticipated in Lenneberg 1967, came back on the agenda of linguists and other cognitive scientists.

We believe that the recent resurgence of interest in 'biolinguistics' is due in large part to the advent of the minimalist program in linguistic theory (Chomsky 1993 and subsequent work). At the heart of the minimalist program is the question of how much of the architecture of the language faculty can be given a principled explanation. Specifically, minimalism asks how well the engine of language meets design requirements imposed by the cognitive systems it subserves. Inevitably, linguists working in the context of the minimalist program are forced to address and sharpen questions of cognitive specificity, ontogeny, phylogeny, and so on, to even begin to understand the design requirements imposed on the language faculty. This is not to say that previous generations of linguists were not interested in such issues. But in practice biolinguistic issues had little effect on empirical inquiry into questions of descriptive and explanatory adequacy.

It is important for us to stress that biolinguistics is independent of the minimalist program. As Lenneberg's work makes clear, biolinguistic questions can be fruitfully addressed outside of a minimalist context. But we think that such a context certainly facilitates, indeed, necessitates inquiry into the biological foundations of language. Last, but not least, we want to remind readers that minimalism is an approach to language that is largely independent of theoretical persuasion. It is an aspect of linguistic research that can be shared by virtually all existing frameworks in linguistic theory that we are familiar with.

3. Our Hope for *Biolinguistics*

To paraphrase Theodosius Dobzhansky's well-known dictum, we think that nothing in language makes sense except in the context of the biology of grammar (cf. Dobzhansky 1973). It is a tribute to Noam Chomsky's own efforts (as well as the efforts of his associates, such as Eric Lenneberg) to treat linguistics as a natural science, and by doing so help her become one, that the term biolinguistics is now seen in course titles, workshops, reading groups, and so on. One can only hope that the term biolinguistics will make its way into institutional categories. Our hope is that this journal will contribute to this exciting and rapidly growing field.

We are fully aware of the fact that the uniquely interdisciplinary character of biolinguistics poses difficult problems of communication and misunderstandings, but we feel that a growing community of scientists of diverse background, including linguists, evolutionary biologists, molecular biologists, neuroscientists, anthropologists, psychologists, computer scientists, (language or speech and hearing) pathologists, and so on, are slowly overcoming these challenges. Only collaboration and mutual respect will make this type of research possible. We would be delighted if the contributions to *Biolinguistics* could clarify issues, unearth new data, and answer some of the questions that will help us understand the nature of language, and what it is that makes us human.

4. Outlook: The First Volume and Beyond

As the journal webpage states, "*Biolinguistics* is a peer-reviewed journal exploring theoretical linguistics that takes the biological foundations of human language seriously" (see http://www.biolinguistics.eu for full text). The high standing of our editorial board members in their respective fields — leading scholars in theoretical linguistics, language acquisition, language change, theoretical biology, genetics, philosophy of mind, and cognitive psychology — helps to ensure a fair and thorough review process. The journal *Biolinguistics* has its own ISSN (1450-3417, as imprinted on every contribution's first page footer as well as back and front cover) and is currently being abstracted and indexed for the usual places. Access to the journal is free, but online user registration is necessary. The full description of the aims, goals, and scope of the journal *Biolinguistics* can be obtained from the website. Subscribers will also receive regular updates and information, and in the near future, interactive tools will be integrated, for which Epstein & Seely's (this volume) multimedia tutorial might just be one example. We encourage submission of products and ideas.

In terms of contributions we accept for submission, *Biolinguistics* features four types:

- *Articles* (full-fledged contributions to the field — complete with abstract, introduction, conclusion — peer-reviewed of ideally 10-12,000 words),
- *Briefs* (very short notes or points, certainly no more than 2,000 words),
- *Reviews* (of recently published books, particular software and other tech equipment, or any other items that warrant a review for *Biolinguistics*), and
- the *Forum* (contributions that don't follow into any of the other categories, such as state-of-the-art reports, research overviews, interviews, and so on).

As can be witnessed, this first volume features all types of contributions: Aside from an editorial (to appear on an irregular basis), it contains four articles (on philosophy, phonology, acquisition, and syntax), one brief (on parameters in acquisition) and one book review (on evolutionary phonology), as well as three forum contributions (a report on experimental syntax, a brief outline for a multimedia tutorial and the relevant link, and an interview).

We would like to close this editorial with an expression of our gratitude to all the people, especially our reviewers and task teams members involved, who helped complete the first volume (see also p. 150 in the "Forum" category at the end of this issue). We would also like to thank the Department of English Studies at the University of Cyprus for substantial financial support.

Appendix: Three Historical Letters

The three letters reproduced here are courtesy of Lyle Jenkins. We would like to thank François Jacob for giving us permission to reprint his letter here.

Institut Pasteur
Département de Biologie Moléculaire
25, RUE DU DOCTEUR ROUX
75724 PARIS Cedex 15
Téléphone : 541-52-66

Paris, le 24 février 1978

Dr. Lyle JENKINS
Institut für Sprachwissenschaft
Liechtensteinstr. 46A/9
A - 1090 - VIENNA

Dear Dr. Jenkins,

Thank you for your letter concerning the Journal "Biolinguistics" you are planning. I am certainly interested in the topics. I would certainly consider participating in the editorial board of the Journal, provided it does not nee too much work, because I am actually near to saturation. I would certainly be pleased to talk with you when you have an occasion of coming to Paris. When you will be in Paris, would you please phone to my laboratory (783-83-12) so that we can manage an appointment.

Yours sincerely,

Jacob

François JACOB

FJ:mc

LINUS PAULING INSTITUTE of SCIENCE and MEDICINE
2700 Sand Hill Road, Menlo Park, California 94025
Telephone: (415) 854-0843

23 April 1980

Dr. Lyle Jenkins
MAXAM, Rm. 830
Division of Cancer Genetics
Sidney Farber Cancer Institute
44 Binney Street
Boston, MA 02115

Dear Dr. Jenkins:

In answer to your letter, I may say that I am pleased to accept your invitation to me to be a member of the editorial board of new journal Biolinguistics.

I mention to you that the President and Director of our Institute is a man who worked with me for a number of years before becoming the founder and director of a CNRS laboratory of macromolecular biology in Montpellier. After ten or eleven years he left that job, and came to the United States and to our Institute. He has been the editor-in-chief of the Journal of Molecular Evolution during the whole eleven years of its existence. You might consider whether or not he should also be a member of your editorial board.

Sincerely,

Linus Pauling

LP:dm

Basel Institute for Immunology

487, Grenzacherstrasse
Postfach
CH-4005 Basel 5, Switzerland

Phone (061) 49 23 80
Cable Antibody Basel
Telex 63458 bii ch

Dr Lyle Jenkins
Maxam, Rm. 830
Division of Genetics
Sidney Farber Cancer Institute
44 Binney Street
BOSTON Massachusetts 02115
USA

15th July, 1980

Dear Dr Jenkins,

Thank you very much for your letter of 5th July concerning your plans to organize a journal and series of books on language and biology. I find this problem of biolinguistics extremely interesting. Linguistic questions have always fascinated me, also in their relation to other areas of biology. Moreover, I am myself multilingual having grown up as a child with three languages (English, Dutch and Danish) and speaking three languages here in Basel (English, French and German). But let me simply answer your question: Yes, I would enjoy joining the Editorial Board of your journal Biolinguistics. I shall look forward to hearing from you how these plans develop.

With kind regards,

Yours sincerely,

Niels Jerne

N. K. Jerne

References

Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms., Harvard University/Massachusetts Institute of Technology. [Published in part as *The Logical Structure of Linguistic Theory*, New York: Plenum, 1975.]

Chomsky, Noam. 1957. *Syntactic Structures* (Januar Linguarum, Series Minor 4). The Hague: Mouton.

Chomsky, Noam. 1959. Review of B.F. Skinner, *Verbal Behavior*. *Language* 35(1), 26-58. [Reprinted in Barbara C. Lust & Claire Foley, eds. 2004. *First Language Acquisition: The Essential Readings* (Linguistics: The Essential Readings), 25-55. Malden, MA: Blackwell.]

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use* (Convergence). New York: Praeger.

Chomsky, Noam. 1988. *Language and Problems of Knowledge: The Managua Lectures* (Current Studies in Linguistics 16). Cambridge, MA: MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In Ken Hale & Samuel Jay Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger* (Current Studies in Linguistics 24), 1-52. Cambridge, MA: MIT Press. [Reprinted in Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28), 167-217. Cambridge, MA: MIT Press.]

Dobzhansky, Theodosius. 1973. Nothing in biology makes sense except in the light of evolution. *The American Biology Teacher* 35(3), 125-129.

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

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

Meader, Clarence L. & John H. Muyskens. 1950. *Handbook of Biolinguistics*. Toledo: H.C. Weller.

Piattelli-Palmarini, Massimo. 1974. A debate on bio-linguistics. *Centre Royaumont pour une science de l'homme report*. [Conference held at Endicott House, Dedham, Massachusetts, 20-21 May, 1974.]

Tinbergen. Nikolaas. 1963. On the aims and methods of ethology. *Zeitschrift für Tierpsychologie* 20(4), 410-433.

Cedric Boeckx
Harvard University
Department of Linguistics
Boylston Hall, 313
Cambridge, MA 02318
USA
cboeckx@fas.harvard.edu

Kleanthes K. Grohmann
University of Cyprus
Department of English Studies
P.O. Box 20537
1087 Nicosia
Cyprus
kleanthi@ucy.ac.cy

Of Minds and Language

Noam Chomsky

This article reviews, and rethinks, a few leading themes of the biolinguistic program since its inception in the early 1950s, at each stage influenced by developments in the biological sciences. The following also discusses how the questions now entering the research agenda develop in a natural way from some of the earliest concerns of these inquiries.

Keywords: biolinguistics; I-language; mind; minimalism

I have been thinking about various ways to approach this opportunity,[1] and on balance, it seemed that the most constructive tack would be to review, and rethink, a few leading themes of the biolinguistic program since its inception in the early 1950s, at each stage influenced by developments in the biological sciences. And to try to indicate how the questions now entering the research agenda develop in a natural way from some of the earliest concerns of these inquiries. Needless to say, this is from a personal perspective. The term "biolinguistics" itself was coined by Massimo Piattelli-Palmarini as the topic for an international conference in 1974 (Piattelli-Palmarini 1974) that brought together evolutionary biologists, neuroscientists, linguists, and others concerned with language and biology, one of many such initiatives, including the Royaumont conference that Massimo brought up (Piattelli-Palmarini 1980).

As you know, the 1950s was the heyday of the behavioral sciences. B.F. Skinner's William James lectures, which later appeared as *Verbal Behavior* (Skinner 1957), were widely circulated by 1950, at least in Cambridge, Mass., and soon became close to orthodoxy, particularly as the ideas were taken up by W.V. Quine in his classes and work that appeared a decade later in his *Word and Object* (Quine 1960). Much the same was assumed for human capacity and cultural variety generally. Zellig Harris's (1951) *Methods of Structural Linguistics* appeared

Editor's note: We are grateful to Noam Chomsky for offering this contribution to the first issue of *Biolinguistics*. This would not have been possible without the support of the editors of the volume in which this article is also going to appear: Massimo Piattelli-Palmarini, Pello Salaburu, and Juan Uriagereka (see fn. 1). We also would like to express our gratitude to Oxford University Press, and to John Davey in particular, for granting us the permission to publish Chomsky's contribution here.

1 Editors' note: The San Sebastian Meeting, June 2006; see Piattelli-Palmarini, Uriagereka & Salaburu (in press).

ISSN 1450–3417
Biolinguistics 1: 009–027, 2007
http://www.biolinguistics.eu

at the same time, outlining procedures for the analysis of a corpus of materials from sound to sentence, reducing data to organized form, and particularly within American linguistics, was generally assumed to have gone about as far as theoretical linguistics could or should reach. The fact that the study was called "methods" reflected the prevailing assumption that there could be nothing much in the way of a theory of language, because languages can "differ from each other without limit and in unpredictable ways," so that the study of each language must be approached "without any preexistent scheme of what a language must be," the formulation of Martin Joos, summarizing the reigning "Boasian tradition," as he plausibly called it. The dominant picture in general biology was in some ways similar, captured in Gunther Stent's (much later) observation that the variability of organisms is so free as to constitute "a near infinitude of particulars which have to be sorted out case by case."

European structuralism was a little different, but not much: Trubetzkoy's *Anleitung*, a classic introduction of phonological analysis (Trubetzkoy 1936, 2001), was similar in conception to the American procedural approaches, and in fact there was very little beyond phonology and morphology, the areas in which languages do appear to differ very widely and in complex ways, a matter of some more general interest, so recent work suggests.

Computers were on the horizon, and it was also commonly assumed that statistical analysis of vast corpora should reveal everything there is to learn about language and its acquisition, a severe misunderstanding of the fundamental issue that has been the primary concern of generative grammar from its origins at about the same time: To determine the structures that underlie semantic and phonetic interpretation of expressions and the principles that enter into growth and development of attainable languages. It was, of course, understood from the early 1950s that as computing power grows, it should ultimately be possible for analysis of vast corpora to produce material that would resemble the data analyzed. Similarly, it would be possible to do the same with videotapes of bees seeking nourishment. The latter might well give better approximations to what bees do than the work of bee scientists, a matter of zero interest to them; they want to discover how it works, resorting to elaborate and ingenious experiments. The former is even more absurd, since it ignores the core problems of the study of language.

A quite separate question is whether various characterizations of the entities and processes of language, and steps in acquisition, might involve statistical analysis and procedural algorithms. That they do was taken for granted in the earliest work in generative grammar, for example, in my *Logical Structure of Linguistic Theory* (*LSLT*, Chomsky 1955). I assumed that identification of chunked word-like elements in phonologically analyzed strings was based on analysis of transitional probabilities — which, surprisingly, turns out to be false, as Thomas Gambell and Charles Yang discovered, unless a simple UG prosodic principle is presupposed. *LSLT* also proposed methods to assign chunked elements to categories, some with an information-theoretic flavor; hand calculations in that pre-computer age had suggestive results in very simple cases, but to my knowledge, the topic has not been further pursued.

Information theory was taken to be a unifying concept for the behavioral sciences, along the lines of Warren Weaver's essay in Shannon & Weaver's (1949/1998) famous monograph. Within the engineering professions, highly influential in these areas, it was a virtual dogma that the properties of language, maybe all human behavior, could be handled within the framework of Markov sources, in fact very elementary ones, not even utilizing the capacity of these simple automata to capture dependencies of arbitrary length. The restriction followed from the general commitment to associative learning, which excluded such dependencies. As an aside, my monograph *Syntactic Structures* (Chomsky 1957) begins with observations on the inadequacy in principle of finite automata, hence Markovian sources, but only because it was essentially notes for courses at MIT, where their adequacy was taken for granted. For similar reasons, the monograph opens by posing the task of distinguishing grammatical from ungrammatical sentences, on the analogy of well-formedness in formal systems, then assumed to be an appropriate model for language. In the much longer and more elaborate unpublished monograph *LSLT* two years earlier (Chomsky 1955), intended only for a few friends, there is no mention of finite automata, and a chapter is devoted to the reasons for rejecting any notion of well-formedness: The task of the theory of language is to generate sound–meaning relations fully, whatever the status of an expression, and in fact much important work then and since has had to do with expressions of intermediate status — the difference, say, between such deviant expressions as (1a) and (1b), that is, empty category principle vs. subjacency violations, still not fully understood.

(1) a. * Which book did they wonder why I wrote?
    b. * Which author did they wonder why wrote that book?

There were some prominent critics, like Karl Lashley, but his very important work on serial order in behavior (Lashley 1951), undermining prevailing associationist assumptions, was unknown, even at Harvard where he was a distinguished professor. Another sign of the tenor of the times.

This is a bit of a caricature, but not much. In fact it is understated, because the prevailing mood was also one of enormous self-confidence that the basic answers had been found, and what remained was to fill in the details in a generally accepted picture.

A few graduate students in the Harvard–MIT complex were skeptics. One was Eric Lenneberg, who went on to found the biology of language; another was Morris Halle. One change over the past 50 years is that we've graduated from sharing a cramped office to being in ample adjacent ones. From the early 1950s, we were reading and discussing work that was then well outside the canon: Lorenz, Tinbergen, Thorpe, and other work in ethology and comparative psychology. Also D'Arcy Thompson (1917/1992), though regrettably we had not come across Turing's work in biology (Turing 1952), and his thesis that "we must envisage a living organism as a special kind of system to which the general laws of physics and chemistry apply […] and because of the prevalence of homologies, we may well suppose, as D'Arcy Thompson has done, that certain physical processes are of very general occurrence." The most recent evaluation of these aspects of Turing's work that I've seen, by Justin Leiber (2001), concludes that

Thompson and Turing "regard teleology, evolutionary phylogeny, natural selection, and history to be largely irrelevant and unfortunately effective distractions from fundamental ahistorical biological explanation," the scientific core of biology. That broad perspective may sound less extreme today after the discovery of master genes, deep homologies, conservation, optimization of neural networks of the kind that Chris Cherniak has demonstrated, and much else, perhaps even restrictions of evolutionary/developmental processes so narrow that "replaying the protein tape of life might be surprisingly repetitive" (quoting a report on feasible mutational paths recently published in *Science*, Weinreich *et al.* 2006, reinterpreting a famous image of Steve Gould's). Another major factor in the development of the biolinguistic perspective was work in recursive function theory and the general theory of computation and algorithms, then just becoming readily available, making it possible to undertake more seriously the inquiry into the formal mechanisms of generative grammars that were being explored from the late 1940s.

These various strands could, it seemed, be woven together to develop a very different approach to problems of language and mind, taking behavior and corpora to be not the object of inquiry, as in the behavioral sciences and structural linguistics, but merely data, and not necessarily the best data, for discovery of the properties of the real object of inquiry: The internal mechanisms that generate linguistic expressions and determine their sound and meaning. The whole system would then be regarded as one of the organs of the body, in this case a cognitive organ, like the systems of planning, interpretation, reflection, and whatever else falls among those aspects of the world loosely "termed mental", which reduce somehow to "the organical structure of the brain". I'm quoting chemist/philosopher Joseph Priestley in the late 18th century, articulating a standard conclusion after Newton had demonstrated, to his great dismay and disbelief, that the world is not a machine, contrary to the core assumptions of the 17th century scientific revolution. It follows that we have no choice but to adopt some non-theological version of what historians of philosophy call "Locke's suggestion": That God might have chosen to "superadd to matter a faculty of thinking" just as he "annexed effects to motion which we can in no way conceive motion able to produce" — notably the property of action at a distance, a revival of occult properties, many leading scientists argued (with Newton's partial agreement).

It is of some interest that all of this seems to have been forgotten. The American Academy of Arts and Sciences published a volume summarizing the results of the Decade of the Brain that ended the 20th century (Mountcastle 1998). The guiding theme, formulated by Vernon Mountcastle, is the thesis of the new biology that "[t]hings mental, indeed minds, are emergent properties of brains, [though] these emergences are […] produced by principles that […] we do not yet understand" (Mountcastle 1998: 1). The same thesis has been put forth in recent years by prominent scientists and philosophers as an "astonishing hypothesis" of the new biology, a "radical" new idea in the philosophy of mind, "the bold assertion that mental phenomena are entirely natural and caused by the neuro-physiological activities of the brain," opening the door to novel and promising inquiries, a rejection of Cartesian mind-body dualism, and so on. All, in fact, reiterate formulations of centuries ago, in virtually the same words, after mind-body dualism became unformulable with the disappearance of the only coherent notion of body (physical, material, etc.) — facts well understood in standard histories of materialism, like Friedrich Lange's (1892) 19th century classic.

It is also of some interest that although the traditional mind-body problem dissolved after Newton, the phrase "mind-body problem" has been resurrected for a problem that is only loosely related to the traditional one. The traditional mind-body problem developed in large part within normal science: Certain phenomena could not be explained by the principles of the mechanical philosophy, the presupposed scientific theory of nature, so a new principle was proposed, some kind of *res cogitans*, a thinking substance, alongside of material substance. The next task would be to discover its properties and to try to unify the two substances. That task was undertaken, but was effectively terminated when Newton undermined the notion of material substance.

What is now called the mind-body problem is quite different. It is not part of normal science. The new version is based on the distinction between the first person and the third person perspective. The first person perspective yields a view of the world presented by one's own experience — what the world looks like, feels like, sounds like to me, and so on. The third person perspective is the picture developed in its most systematic form in scientific inquiry, which seeks to understand the world from outside any particular personal perspective.

The new version of the mind-body problem resurrects a thought experiment of Bertrand Russell's 80 years ago, though the basic observation traces back to the pre-Socratics. Russell asked us to consider a blind physicist who knows all of physics but doesn't know something we know: What it's like to see the color blue: "It is obvious that a man who can see knows things which a blind man cannot know; but a blind man can know the whole of physics. Thus the knowledge which other men have and he has not is not part of physics" (Russell 2003: 227). Russell's conclusion was that the natural sciences seek to discover "the causal skeleton of the world. Other aspects lie beyond their purview" (*ibid.*).

Recasting Russell's experiment in naturalistic terms, we might say that like all animals, our internal cognitive capacities reflexively provide us with a world of experience — the human *Umwelt*, in ethological lingo. But being reflective creatures, thanks to emergence of human intellectual capacities, we go on to seek a deeper understanding of the phenomena of experience. If humans are part of the organic world, we expect that our capacities of understanding and explanation have fixed scope and limits, like any other natural object — a truism that is sometimes thoughtlessly derided as "mysterianism," though it was understood by Descartes and Hume, among others. It could be that these innate capacities do not lead us beyond some theoretical understanding of Russell's causal skeleton of the world. In principle these questions are subject to empirical inquiry into what we might call "the science-forming faculty," another "mental organ," now the topic of some investigation — Susan Carey's work, for example. But these issues are distinct from traditional dualism, which evaporated after Newton.

This is a rough sketch of the intellectual background of the biolinguistic perspective, in part with the benefit of some hindsight. Adopting this perspective, the term "language" means internal language, a state of the computational system of the mind/brain that generates structured expressions, each of which can be taken to be a set of instructions for the interface systems within which the faculty of language is embedded. There are at least two such interfaces: The systems of thought that use linguistic expressions for reasoning, interpretation, organizing action, and other mental acts. And the sensorimotor systems that externalize expressions in production and construct them from sensory data in perception. The theory of the genetic endowment for language is commonly called universal grammar (UG), adapting a traditional term to a different framework. Certain configurations are possible human languages, others are not, and a primary concern of the theory of human language is to establish the distinction between the two categories.

Within the biolinguistic framework, several tasks immediately arise. The first is to construct generative grammars for particular languages that yield the facts about sound and meaning. It was quickly learned that the task is formidable. Very little was known about languages, despite millennia of inquiry. The most extensive existing grammars and dictionaries were, basically, lists of examples and exceptions, with some weak generalizations. It was assumed that anything beyond could be determined by unspecified methods of "analogy" or "induction" or "habit." But even the earliest efforts revealed that these notions concealed vast obscurity. Traditional grammars and dictionaries tacitly appeal to the understanding of the reader, either knowledge of the language in question or the shared innate linguistic capacity, or commonly both. But for the study of language as part of biology, it is precisely that presupposed understanding that is the topic of investigation, and as soon as the issue was faced, major problems were quickly unearthed.

The second task is to account for the acquisition of language, later called the problem of explanatory adequacy (when viewed abstractly). In biolinguistic terms, that means discovering the operations that map presented data to the internal language attained. With sufficient progress in approaching explanatory adequacy, a further and deeper task comes to the fore: To transcend explanatory adequacy, asking not just what the mapping principles are, but why language growth is determined by these principles, not innumerable others that can be easily imagined. The question was premature until quite recently, when it has been addressed in what has come to be called the minimalist program, the natural next stage of biolinguistic inquiry, to which I'll briefly return.

Another question is how the faculty of language evolved. There are libraries of books and articles about evolution of language — in rather striking contrast to the literature, say, on the evolution of the communication system of bees. For human language, the problem is vastly more difficult for obvious reasons, and can be undertaken seriously, by definition, only to the extent that some relatively firm conception of UG is available, since that is what evolved.

Still another question is how the properties "termed mental" relate to "the organical structure of the brain," in Priestley's words (see also Chomsky 1998). And there are hard and important questions about how the internal language is put to use, for example in acts of referring to the world, or in interchange with others, the topic of interesting work in neo-Gricean pragmatics in recent years.

Other cognitive organs can perhaps be studied along similar lines. In the early days of the biolinguistic program, George Miller and others sought to construct a generative theory of planning, modeled on early ideas about generative grammar (Miller & Johnson-Laird 1976). Other lines of inquiry trace back to David Hume, who recognized that knowledge and belief are grounded in a "species of natural instincts," part of the "springs and origins" of our inherent mental nature, and that something similar must be true in the domain of moral judgment. The reason is that our moral judgments are unbounded in scope and that we constantly apply them in systematic ways to new circumstances. Hence they too must be founded on general principles that are part of our nature though beyond our "original instincts," those shared with animals. That should lead to efforts to develop something like a grammar of moral judgment. That task was undertaken by John Rawls, who adapted models of generative grammar that were being developed as he was writing his classic *Theory of Justice* (1971) in the 1960s. These ideas have recently been revived and developed and have become a lively field of theoretical and empirical inquiry (cf. Hauser 2006).

At the time of the 1974 biolinguistics conference, it seemed that the language faculty must be rich, highly structured, and substantially unique to this cognitive system. In particular, that conclusion followed from considerations of language acquisition. The only plausible idea seemed to be that language acquisition is rather like theory construction. Somehow, the child reflexively categorizes certain sensory data as linguistic experience, and then uses the experience as evidence to construct an internal language — a kind of theory of expressions that enter into the myriad varieties of language use.

To give a few of the early illustrations for concreteness, the internal language that we more or less share determines that sentence (2a) is three-ways ambiguous, though it may take a little reflection to reveal the fact; but the ambiguities are resolved if we ask (2b), understood approximately as (2c).

(2) a. Mary saw the man leaving the store.
b. Which store did Mary see the man leaving?
c. Which store did Mary see the man leave?

The phrase *which store* is raised from the position in which its semantic role is determined as object of *leave*, and is then given an additional interpretation as an operator taking scope over a variable in its original position, so the sentence means, roughly, *for which x, x a store, Mary saw the man leav(ing) the store x* — and without going into it here, there is good reason to suppose that the semantic interface really does interpret the variable x as *the store x*, a well-studied phenomenon called "reconstruction". The phrase that serves as the restricted variable is silent in the phonetic output, but must be there for interpretation. Only one of the underlying structures permits the operation, so the ambiguity is resolved in the interrogative, in the manner indicated. The constraints involved — so-called "island conditions" — have been studied intensively for about 45 years. Recent work indicates that they may reduce in large measure to minimal search conditions of optimal computation, perhaps not coded in UG but more general laws of nature — which, if true, would carry us beyond explanatory adequacy.

Note that even such elementary examples as this illustrate the marginal interest of the notions "well-formed" or "grammatical" or "good approximation to a corpus", however they are characterized.

To take a second example, illustrating the same principles less transparently, consider sentences (3a) and (3b).

(3) a. John ate an apple.
    b. John ate.

We can omit *an apple*, yielding (3b), which we understand to mean *John ate something unspecified*. Now consider

(4) a. John is too angry to eat an apple.
    b. John is too angry to eat.

We can omit *an apple*, yielding (4b), which, by analogy to (3b) should mean that *John is so angry that he wouldn't eat anything*. That's a natural interpretation, but there is also a different one in this case: namely, *John is so angry that someone or other won't eat him, John* — the natural interpretation for the structurally analogous expression

(5) John is too angry to invite.

In this case, the explanation lies in the fact that the phrase *too angry to eat* does include the object of *eat*, but it is invisible. The invisible object is raised just as *which store* is raised in the previous example (2), again yielding an operator-variable structure. In this case, however, the operator has no content, so the construction is an open sentence with a free variable, hence a predicate. The semantic interpretation follows from general principles. The minimal search conditions that restrict raising of *which store* in example (2) also bar the raising of the empty object of *eat*, yielding standard island properties.

In both cases, the same general computational principles, operating efficiently, provide a specific range of interpretations as an operator-variable construction, with the variable unpronounced in both cases and the operator unpronounced in one. The surface forms in themselves tell us little about the interpretations.

Even the most elementary considerations yield the same conclusions. The simplest lexical items raise hard if not insuperable problems for analytic procedures of segmentation, classification, statistical analysis, and the like. A lexical item is identified by phonological elements that determine its sound along with morphological elements that determine its meaning. But neither the phonological nor morphological elements have the "beads-on-a-string" property required for computational analysis of a corpus. Furthermore, even the simplest words in many languages have phonological and morphological elements that are silent. The elements that constitute lexical items find their place in the generative procedures that yield the expressions, but cannot be detected in the physical signal. For that reason, it seemed then — and still seems — that the language acquired must have the basic properties of an internalized explanatory theory. These are design properties that any account of evolution of language must deal with.

Quite generally, construction of theories must be guided by what Charles Sanders Peirce a century ago called an "abductive principle," which he took to be a genetically determined instinct, like the pecking of a chicken. The abductive principle "puts a limit upon admissible hypotheses" so that the mind is capable of "imagining correct theories of some kind" and discarding infinitely many others consistent with the evidence. Peirce was concerned with what I was calling "the science-forming faculty," but similar problems arise for language acquisition, though it is dramatically unlike scientific discovery. It is rapid, virtually reflexive, convergent among individuals, relying not on controlled experiment or instruction but only on the "blooming, buzzing confusion" that each infant confronts. The format that limits admissible hypotheses about structure, generation, sound and meaning must therefore be highly restrictive. The conclusions about the specificity and richness of the language faculty follow directly. Plainly such conclusions make it next to impossible to raise questions that go beyond explanatory adequacy — the "why" questions — and also pose serious barriers to inquiry into how the faculty might have evolved, matters discussed inconclusively at the 1974 conference (see Piattelli-Palmarini 1974).

A few years later, a new approach suggested ways in which these paradoxes might be overcome. This Principles–and–Parameters (P&P) approach (Chomsky 1981 *et seq.*) was based on the idea that the format consists of invariant principles and a "switch-box" of parameters — to adopt Jim Higginbotham's image. The switches can be set to one or another value on the basis of fairly elementary experience. A choice of parameter settings determines a language. The approach largely emerged from intensive study of a range of languages, but as in the early days of generative grammar, it was also suggested by developments in biology — in this case, François Jacob's ideas about how slight changes in the timing and hierarchy of regulatory mechanisms might yield great superficial differences (a butterfly or an elephant, and so on). The model seemed natural for language as well: Slight changes in parameter settings might yield superficial variety, through interaction of invariant principles with parameter choices. That's discussed a bit in Kant lectures of mine at Stanford in 1978, which appeared a few years later in my book *Rules and Representations* (Chomsky 1980).

The approach crystallized in the early 1980s, and has been pursued with considerable success, with many revisions and improvements along the way. One illustration is Mark Baker's demonstration, in his book *Atoms of Language* (Baker 2001), that languages that appear on the surface to be about as different as can be imagined (in his case Mohawk and English) turn out to be remarkably similar when we abstract from the effects of a few choices of values for parameters within a hierarchic organization that he argues to be universal, hence the outcome of evolution of language.

Looking with a broader sweep, the problem of reconciling unity and diversity has constantly arisen in biology and linguistics. The linguistics of the early scientific revolution distinguished universal from particular grammar, though not in the biolinguistic sense. Universal grammar was taken to be the intellectual core of the discipline; particular grammars are accidental instantiations. With the flourishing of anthropological linguistics, the pendulum swung in the other direction, towards diversity, well captured in the Boasian formulation to which I referred. In general biology, a similar issue had been raised sharply in the Cuvier–Geoffroy debate in 1830 (Appel 1987). Cuvier's position, emphasizing diversity, prevailed, particularly after the Darwinian revolution, leading to the conclusions about near infinitude of variety that have to be sorted out case by case, which I mentioned earlier. Perhaps the most quoted sentence in biology is Darwin's final observation in *Origin of Species* about how "from so simple a beginning, endless forms most beautiful and most wonderful have been, and are being, evolved." I don't know if the irony was intended, but these words were taken by Sean Carroll (2005) as the title of his introduction to *The New Science of Evo Devo*, which seeks to show that the forms that have evolved are far from endless, in fact are remarkably uniform, presumably, in important respects, because of factors of the kind that Thompson and Turing thought should constitute the true science of biology. The uniformity had not passed unnoticed in Darwin's day. Thomas Huxley's naturalistic studies led him to observe that there appear to be "predetermined lines of modification" that lead natural selection to "produce varieties of a limited number and kind" for each species.[2]

Over the years, in both general biology and linguistics the pendulum has been swinging towards unity, in the evo–devo revolution in biology and in the somewhat parallel minimalist program.

The principles of traditional universal grammar had something of the status of Joseph Greenberg's universals: They were descriptive generalizations. Within the framework of UG in the contemporary sense, they are observations to be explained by the principles that enter into generative theories, which can be investigated in many other ways. Diversity of language provides an upper bound on what may be attributed to UG: It cannot be so restricted as to exclude attested languages. Poverty of stimulus (POS) considerations provide a lower bound: UG must be at least rich enough to account for the fact that internal languages are attained. POS considerations were first studied seriously by Descartes to my knowledge, in the field of visual perception. Of course they are central to any inquiry into growth and development, though for curious reasons, these truisms are considered controversial only in the case of language and other higher human mental faculties (particular empirical assumptions about POS are of

---

[2] The passage quoted is, in its entirety:

> The importance of natural selection will not be impaired even if further inquiries should prove that variability is definite, and is determined in certain directions rather than in others, by conditions inherent in that which varies. It is quite conceivable that every species tends to produce varieties of a limited number and kind, and that the effect of natural selection is to favour the development of some of these, while it opposes the development of others along their predetermined lines of modification. (Huxley 1893: 223)

See also Gates (1916: 128) and Chomsky (2004).

course not truisms, in any domain of growth and development).

For these and many other reasons, the inquiry has more stringent conditions to satisfy than generalization from observed diversity. That is one of many consequences of the shift to the biolinguistic perspective; another is that methodological questions about simplicity, redundancy, and so on, are transmuted into factual questions that can be investigated from comparative and other perspectives, and may reduce to natural law.

Apart from stimulating highly productive investigation of languages of great typological variety, at a depth never before even considered, the P&P approach also reinvigorated neighboring fields, particularly the study of language acquisition, reframed as inquiry into setting of parameters in the early years of life. The shift of perspective led to very fruitful results, enough to suggest that the basic contours of an answer to the problems of explanatory adequacy might be visible. On that tentative assumption, we can turn more seriously to the "why" questions that transcend explanatory adequacy. The minimalist program thus arose in a natural way from the successes of the P&P approach.

The P&P approach also removed the major conceptual barrier to the study of evolution of language. With the divorce of principles of language from acquisition, it no longer follows that the format that "limits admissible hypotheses" must be rich and highly structured to satisfy the empirical conditions of language acquisition, in which case inquiry into evolution would be virtually hopeless. That might turn out to be the case, but it is no longer an apparent conceptual necessity. It therefore became possible to entertain more seriously the recognition, from the earliest days of generative grammar, that acquisition of language involves not just a few years of experience and millions of years of evolution, yielding the genetic endowment, but also "principles of neural organization that may be even more deeply grounded in physical law" (quoting from my *Aspects of the Theory of Syntax*, Chomsky 1965 — a question then premature).

Assuming that language has general properties of other biological systems, we should be seeking three factors that enter into its growth in the individual: (i) genetic factors, the topic of UG, (ii) experience, which permits variation within a fairly narrow range, and (iii) principles not specific to language. The third factor includes principles of efficient computation, which would be expected to be of particular significance for systems such as language. UG is the residue when third factor effects are abstracted. The richer the residue, the harder it will be to account for the evolution of UG, evidently.

Throughout the modern history of generative grammar, the problem of determining the general nature of language has been approached "from top down," so to speak: How much must be attributed to UG to account for language acquisition? The minimalist program seeks to approach the problem "from bottom up": How little can be attributed to UG while still accounting for the variety of internal languages attained, relying on third factor principles? Let me end with a few words on this approach.

An elementary fact about the language faculty is that it is a system of discrete infinity. In the simplest case, such a system is based on a primitive operation that takes objects already constructed, and constructs from them a new object. Call that operation Merge. There are more complex modes of generation, such as the familiar phrase structure grammars explored in the early years of generative grammar. But a Merge-based system is the most elementary, so we assume it to be true of language unless empirical facts force greater UG complexity. If computation is efficient, then when X and Y are merged, neither will change, so that the outcome can be taken to be simply the set {X,Y}. That is sometimes called the No-Tampering condition, a natural principle of efficient computation, perhaps a special case of laws of nature. With Merge available, we instantly have an unbounded system of hierarchically structured expressions. For language to be usable, these expressions have to link to the interfaces. The generated expressions provide the means to relate sound and meaning in traditional terms, a far more subtle process than had been assumed for millennia. UG must at least include the principle of unbounded Merge.

The conclusion holds whether recursive generation is unique to the language faculty or found elsewhere. If the latter, there still must be a genetic instruction to use unbounded Merge to form linguistic expressions. Nonetheless, it is interesting to ask whether this operation is language-specific. We know that it is not. The classic illustration is the system of natural numbers, raising problems for evolutionary theory noted by Alfred Russel Wallace. A possible solution is that the number system is derivative from language. If the lexicon is reduced to a single element, then unbounded Merge will easily yield arithmetic. Speculations about the origin of the mathematical capacity as an abstraction from language are familiar, as are criticisms, including apparent dissociation with lesions and diversity of localization. The significance of such phenomena, however, is far from clear. As Luigi Rizzi has pointed out (Rizzi 2003), they relate to use of the capacity, not its possession; for similar reasons, dissociations do not show that the capacity to read is not parasitic on the language faculty. The competence-performance distinction should not be obscured. To date, I am not aware of any real examples of unbounded Merge apart from language, or obvious derivatives from language, for example, taking visual arrays as lexical items.

We can regard an account of some linguistic phenomena as principled insofar as it derives them by efficient computation satisfying interface conditions. A very strong proposal, called "the strong minimalist thesis," is that all phenomena of language have a principled account in this sense, that language is a perfect solution to interface conditions, the conditions it must satisfy to some extent if it is to be usable at all. If that thesis were true, language would be something like a snowflake, taking the form it does by virtue of natural law, in which case UG would be very limited.

In addition to unbounded Merge, language requires atoms, or word-like elements, for computation. Whether these belong strictly to language or are appropriated from other cognitive systems, they pose extremely serious problems for the study of language and thought and also for the study of the evolution of human cognitive capacities. The basic problem is that even the simplest words and concepts of human language and thought lack the relation to mind-independent entities that has been reported for animal communication:

Representational systems based on a one-to-one relation between mind/brain processes and "an aspect of the environment to which these processes adapt the animal's behavior," to quote Randy Gallistel. The symbols of human language and thought are sharply different.

These matters were explored in interesting ways by 17th-18th century British philosophers, developing ideas that trace back to Aristotle. Carrying their work further, we find that human language appears to have no reference relation, in the sense stipulated in the study of formal systems, and presupposed — mistakenly, I think — in contemporary theories of reference for language in philosophy and psychology, which take for granted some kind of word-object relation, where the objects are extra-mental. What we understand to be a house, a river, a person, a tree, water, and so on, consistently turns out to be a creation of what 17th century investigators called the "cognoscitive powers," which provide us with rich means to refer to the outside world from certain perspectives. The objects of thought they construct are individuated by mental operations that cannot be reduced to a "peculiar nature belonging" to the thing we are talking about, as David Hume summarized a century of inquiry. There need be no mind-independent entity to which these objects of thought bear some relation akin to reference, and apparently there is none in many simple cases (probably all). In this regard, internal conceptual symbols are like the phonetic units of mental representations, such as the syllable /ba/; every particular act externalizing this mental entity yields a mind-independent entity, but it is idle to seek a mind-independent construct that corresponds to the syllable. Communication is not a matter of producing some mind-external entity that the hearer picks out of the world, the way a physicist could. Rather, communication is a more-or-less affair, in which the speaker produces external events and hearers seek to match them as best they can to their own internal resources. Words and concepts appear to be similar in this regard, even the simplest of them. Communication relies on shared cognoscitive powers, and succeeds insofar as shared mental constructs, background, concerns, presuppositions, etc. allow for common perspectives to be (more or less) attained. These semantic properties of lexical items seem to be unique to human language and thought, and have to be accounted for somehow in the study of their evolution.

Returning to the computational system, as a simple matter of logic, there are two kinds of Merge, external and internal. External Merge takes two objects, say *eat* and *apples*, and forms the new object that corresponds to *eat apples*. Internal Merge — often called Move — is the same, except that one of the objects is internal to the other. So applying internal Merge to *John ate what*, we form the new object corresponding to *what John ate what*, in accord with the No-Tampering condition. As in the examples I mentioned earlier, at the semantic interface, both occurrences of *what* are interpreted: The first occurrence as an operator and the second as the variable over which it ranges, so that the expression means something like *for which thing x, John ate the thing x*. At the sensorimotor side, only one of the two identical syntactic objects is pronounced, typically the structurally most salient occurrence. That illustrates the ubiquitous displacement property of language: Items are commonly pronounced in one position but interpreted somewhere else as well. Failure to pronounce all but one occurrence follows from third factor considerations of efficient computation, since it reduces the burden of repeated application of the rules that transform internal structures to phonetic form — a heavy burden when we consider real cases. There is more to say, but this seems the heart of the matter.

This simple example suggests that the relation of the internal language to the interfaces is asymmetrical. Optimal design yields the right properties at the semantic side, but causes processing problems at the sound side. To understand the perceived sentence (6),

(6) What did John eat?

it is necessary to locate and fill in the missing element, a severe burden on speech perception in more complex constructions. Here conditions of efficient computation conflict with facilitation of communication. Universally, languages prefer efficient computation. That appears to be true more generally. For example, island conditions are at least sometimes, and perhaps always, imposed by principles of efficient computation. They make certain thoughts inexpressible, except by circumlocution, thus impeding communication. The same is true of ambiguities, as in the examples I mentioned earlier. Structural ambiguities often fall out naturally from efficient computation, but evidently pose a communication burden.

Other considerations suggest the same conclusion. Mapping to the sensorimotor interface appears to be a secondary process, relating systems that are independent: the sensorimotor system, with its own properties, and the computational system that generates the semantic interface, optimally insofar as the strong minimalist thesis is accurate. That's basically what we find. Complexity, variety, effects of historical accident, and so on, are overwhelmingly restricted to morphology and phonology, the mapping to the sensorimotor interface. That's why these are virtually the only topics investigated in traditional linguistics, or that enter into language teaching. They are idiosyncrasies, so are noticed, and have to be learned. If so, then it appears that language evolved, and is designed, primarily as an instrument of thought. Emergence of unbounded Merge in human evolutionary history provides what has been called a "language of thought," an internal generative system that constructs thoughts of arbitrary richness and complexity, exploiting conceptual resources that are already available or may develop with the availability of structured expressions. If the relation to the interfaces is asymmetric, as seems to be the case, then unbounded Merge provides only a language of thought, and the basis for ancillary processes of externalization.

There are other reasons to believe that something like that is true. One is that externalization appears to be independent of sensory modality, as has been learned from studies of sign language in recent years. More general considerations suggest the same conclusion. The core principle of language, unbounded Merge, must have arisen from some rewiring of the brain, presumably the effect of some small mutation. Such changes take place in an individual, not a group. The individual so endowed would have had many advantages: capacities for complex thought, planning, interpretation, and so on.

The capacity would be transmitted to offspring, coming to dominate a small breeding group. At that stage, there would be an advantage to externalization, so the capacity would be linked as a secondary process to the sensorimotor system for externalization and interaction, including communication. It is not easy to imagine an account of human evolution that does not assume at least this much. And empirical evidence is needed for any additional assumption about the evolution of language.

Such evidence is not easy to find. It is generally supposed that there are precursors to language proceeding from single words, to simple sentences, then more complex ones, and finally leading to unbounded generation. But there is no empirical evidence for the postulated precursors, and no persuasive conceptual argument for them either: Transition from 10-word sentences to unbounded Merge is no easier than transition from single words. A similar issue arises in language acquisition. The modern study of the topic began with the assumption that the child passes through a one and two-word stage, telegraphic speech, and so on. Again the assumption lacks a rationale, because at some point unbounded Merge must appear. Hence the capacity must have been there all along even if it only comes to function at some later stage. There does appear to be evidence about earlier stages: namely, what children produce. But that carries little weight. Children understand far more than what they produce, and understand normal language but not their own restricted speech, as was shown long ago by Lila Gleitman and her colleagues (Shipley *et al.* 1969). For both evolution and development, there seems little reason to postulate precursors to unbounded Merge.

In the 1974 biolinguistics conference, evolutionary biologist Salvador Luria was the most forceful advocate of the view that communicative needs would not have provided "any great selective pressure to produce a system such as language," with its crucial relation to "development of abstract or productive thinking." His fellow Nobel laureate François Jacob (1977) added later that "the role of language as a communication system between individuals would have come about only secondarily, as many linguists believe," perhaps referring to discussions at the symposia (for an insightful reconstruction of those debates, see also Jenkins 2000). "The quality of language that makes it unique does not seem to be so much its role in communicating directives for action" or other common features of animal communication, Jacob continues, but rather "its role in symbolizing, in evoking cognitive images," in "molding" our notion of reality and yielding our capacity for thought and planning, through its unique property of allowing "infinite combinations of symbols" and therefore "mental creation of possible worlds," ideas that trace back to the 17th century cognitive revolution and have been considerably sharpened in recent years.

We can, however, go beyond speculation. Investigation of language design can yield evidence on the relation of language to the interfaces. There is, I think, mounting evidence that the relation is asymmetrical in the manner indicated. There are more radical proposals under which optimal satisfaction of semantic conditions becomes close to tautologous. That seems to me one way to understand the general drift of Jim Higginbotham's work on the syntax-semantics border for many years. And from a different point of view, something similar would follow from ideas developed by Wolfram Hinzen (2006, 2007), in line with Juan Uriagereka's suggestion that it is "as if syntax carved the path interpretation must blindly follow."

The general conclusions appear to fit reasonably well with evidence from other sources. It seems that brain size reached its current level about 100,000 years ago, which suggests to some specialists that "human language probably evolved, at least in part, as an automatic but adaptive consequence of increased absolute brain size," leading to dramatic changes of behavior (quoting George Striedter, in *Brain and Behavioral Sciences* February 2006, who adds qualifications about the structural and functional properties of primate brains; Striedter 2006: 9). This "great leap forward," as some call it, must have taken place before about 50,000 years ago, when the trek from Africa began. Even if further inquiry extends the boundaries, it remains a small window, in evolutionary time. The picture is consistent with the idea that some small rewiring of the brain gave rise to unbounded Merge, yielding a language of thought, later externalized and used in many ways. Aspects of the computational system that do not yield to principled explanation fall under UG, to be explained somehow in other terms, questions that may lie beyond the reach of contemporary inquiry, Richard Lewontin (1998) has argued. Also remaining to be accounted for are the apparently human-specific atoms of computation, the minimal word-like elements of thought and language, and the array and structure of parameters, rich topics that I barely mentioned.

At this point we have to move on to more technical discussion than is possible here, but I think it is fair to say that there has been considerable progress in moving towards principled explanation in terms of third factor considerations. The best guess about the nature of UG only a few years ago has been substantially improved by approaching the topic "from bottom up", by asking how far we can press the strong minimalist thesis. It seems now that much of the architecture that has been postulated can be eliminated without loss, often with empirical gain. That includes the last residues of phrase structure grammar, including the notion of projection or later "labeling," the latter perhaps eliminable in terms of minimal search. Also eliminable on principled grounds are underlying and surface structure, and also logical form, in its technical sense, leaving just the interface levels (and their existence too is not graven in stone, a separate topic). The several compositional cycles that have commonly been postulated can be reduced to one, with periodic transfer of generated structures to the interface at a few designated positions ("phases"), yielding further consequences. A very elementary form of transformational grammar essentially "comes free;" it would require stipulations to block it, so that there is a principled explanation, in these terms, for the curious but ubiquitous phenomenon of displacement in natural language, with interpretive options in positions that are phonetically silent. And by the same token, any other approach to the phenomenon carries an empirical burden. Some of the island conditions have principled explanations, as does the existence of categories for which there is no direct surface evidence, such as a functional category of inflection.

Without proceeding, it seems to me no longer absurd to speculate that there may be a single internal language, efficiently yielding the infinite array of expressions that provide a language of thought. Variety and complexity of language would then be reduced to the lexicon, which is also the locus of parametric variation, and to the ancillary mappings involved in externalization, which might turn out to be best possible solutions to relating organs with independent origins and properties. There are huge promissory notes left to pay, and alternatives that merit careful consideration, but plausible reduction of the previously assumed richness of UG has been substantial.

With each step towards the goals of principled explanation we gain a clearer grasp of the essential nature of language, and of what remains to be explained in other terms. It should be kept in mind, however, that any such progress still leaves unresolved problems that have been raised for hundreds of years. Among these is the question how properties "termed mental" relate to "the organical structure of the brain," in the 18th century formulation. And beyond that lies the mysterious problem of the creative and coherent ordinary use of language, a central problem of Cartesian science, still scarcely even at the horizons of inquiry.

References

Appel, Toby A. 1987. *The Cuvier–Geoffroy Debate: French Biology in the Decades before Darwin* (Monographs on the History and Philosophy of Biology). New York: Oxford University Press.

Baker, Mark. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.

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

Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms., Harvard University/Massachusetts Institute of Technology. [Published in part as *The Logical Structure of Linguistic Theory*, New York: Plenum, 1975.]

Chomsky, Noam. 1957. *Syntactic Structures* (Januar Linguarum, Series Minor 4). The Hague: Mouton.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1980. *Rules and Representations* (Woodbridge Lectures 11). New York: Columbia University Press.

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

Chomsky, Noam. 1998. Comments: Galen Strawson, "Mental Reality". *Philosophy and Phenomenological Research* 58(2), 437-441.

Chomsky, Noam. 2004. The biolinguistic perspective after 50 years. *Quaderni del Dipartimento di Linguistica, Universita' di Firenze* 14, 3-12. [Also published in *Sito Web dell' Accademia della Crusca*, 2005. http://www.accademiadellacrusca.it/img_usr/Chomsky%20-%20ENG.pdf (19 November 2007).]

Darwin, Charles. 1892. *The Origin of Species*. London: John Murray.

Gates, R. Ruggles. 1916. Huxley as mutationist. *The American Naturalist* 50(592), 126-128.
Harris, Zellig S. 1951. *Methods of Structural Linguistics*. Chicago: University of Chicago Press.
Hauser, Marc D. 2006. *Moral Minds: How Nature Designed Our Universal Sense of Right and Wrong*. New York: Ecco.
Hinzen, Wolfram. 2006. *Mind Design and Minimal Syntax*. Oxford: Oxford University Press.
Hinzen, Wolfram. 2007. *An Essay on Names and Truth*. Oxford: Oxford University Press.
Huxley, Thomas. 1893. *Darwiniana: Essays*. London: Macmillan.
Jacob, François. 1977. Evolution and tinkering. *Science* 196(1), 161-171.
Jenkins, Lyle. 2000. *Biolinguistics: Exploring the Biology of Language*. Cambridge: Cambridge University Press.
Lange, Friedrich A. 1892. *History of Materialism and Criticism of Its Present Importance*, vol. 3 [transl. by Ernest Chester Thomas]. London: Kegan Paul, Trench, Trübner & Co. [3rd edn., with an introduction by Bertrand Russell, New York: The Humanities Press, reprinted 1950.]
Lashley, Karl S. 1951. The problem of serial order in behavior. In Lloyd A. Jeffress (ed.), *Cerebral Mechanisms in Behavior*, 112-136. New York: John Wiley.
Leiber, Justin. 2001. Turing and the fragility and insubstantiality of evolutionary explanations: A puzzle about the unity of Alan Turing's work with some larger implications. *Philosophical Philosophy* 14(1), 83-94.
Lewontin, Richard C. 1998. The evolution of cognition: Questions we will never answer. In Don Scarborough & Saul Sternberg (eds.), *Methods, Models, and Conceptual Issues* (An Invitation to Cognitive Science 4, 2nd edn.), 107-134. Cambridge, MA: MIT Press.
Miller, George A. & Philip N. Johnson-Laird. 1976. *Language and Perception*. Cambridge: Cambridge University Press.
Mountcastle, Vernon B. 1998. Brain science at the century's ebb. *Daedalus* 127(2), 1-36.
Piattelli-Palmarini, Massimo. 1974. A debate on bio-linguistics. *Centre Royaumont pour une science de l'homme report*. [Conference held at Endicott House, Dedham, Massachusetts, 20-21 May, 1974.]
Piattelli-Palmarini, Massimo (ed.). 1980. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. London: Routledge & Kegan Paul.
Piattelli-Palmarini, Massimo, Juan Uriagereka & Pello Salaburu (eds.). In press. *Of Minds and Language: The Basque Country Encounter with Noam Chomsky*. New York: Oxford University Press.
Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rizzi, Luigi. 2003. Some elements of the study of language as a cognitive capacity. In Nicola Dimitri (ed.), *Cognitive Processes and Economic Behavior* (Routledge Siena Studies in Political Economy). London: Routledge.
Russell, Bertrand. 2003. *Russell on Metaphysics: Selections from the Writings of Bertrand Russell* [ed. by Stephen Mumford]. London: Routledge.
Quine, W.V.O. 1960. *Word and Object* (Studies in Communication). Cambridge, MA: MIT Press.

Shannon, Claude E. & Warren Weaver. 1949/1998. *The Mathematical Theory of Communication*. Urbana, IL: University of Illinois Press.

Shipley, Elizabeth P., Carlota S. Smith & Lila R. Gleitman. 1969. A study in the acquisition of language: Free responses to commands. *Language* 45 (2), 322-42.

Skinner, B.F. 1957. *Verbal Behavior*. New York: Appleton–Century–Crofts.

Striedter, George F. 2006. Précis of *Principles of Brain Evolution*. *Behavioral and Brain Sciences* 29(1), 1-12.

Thompson, D'Arcy Wentworth. 1917/1992. *On Growth and Form* [abridged edn., prepared by John Tyler Bonner]. Cambridge: Cambridge University Press.

Trubetzkoy, Nicolaj S. 1936. *Anleitung zu phonologischen Beschreibungen*. Brno: Edition du Cercle Linguistique de Prague.

Trubetzkoy, Nicolaj S. 2001. *Studies in General Linguistics and Language Structure* [ed. by Anatoly Liberman, transl. by Marvin Taylor & Anatoly Liberman]. Durham, NC: Duke University Press.

Turing, Alan M. 1952. The chemical basis of morphogenesis. *Philosophical Transactions of the Royal Society of London* 237(B 641), 37-72. [Reprinted in Turing, Alan M. 1992. *Morphogenesis* [ed. by Peter T. Saunders]. Amsterdam: Elsevier.]

Weinreich, Daniel M., Nigel F. Delaney, Mark A. DePristo & Daniel L. Hartl. 2006. Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312(5770), 111-114.

Noam Chomsky
Massachusetts Institute of Technology
Department of Linguistics and Philosophy
Mass. Avenue 32–D808
Cambridge, MA 02139
USA
chomsky@mit.edu

Computing Long-Distance Dependencies in Vowel Harmony

Frédéric Mailhot & Charles Reiss

This article develops an explicit procedural model of vowel harmony, and takes steps toward finding a lower bound on the computational power of phonological rules. The focus on formalization and procedural computation allows for simplification in models of representation and the discovery of interesting interactions involving the conditions in rules. It is shown that locality principles are derivable, which motivates the elimination of iterative rule application advocated here. Along the way, a novel analysis of neutral vowels in harmony processes is also provided.

Keywords: assimilation; locality; long-distance dependency; procedural phonology; SEARCH and COPY; substance-free; vowel harmony

1. Introduction

This article introduces the following proposals concerning the formal properties of a strongly procedural model of phonological assimilation:

(1) *Phonological Assimilation*

a. Assimilatory processes comprise (i) a SEARCH algorithm from which locality effects can be derived (i.e. there are no locality principles encoded in the grammar) and (ii) a COPY operation which transmits feature values across segments.

b. SEARCH-derived locality relations are non-symmetric: If x is in a locality relation $L(x,y)$ with y, it is not necessarily the case that y is in a locality relation $L(y,x)$ with x.

c. SEARCH is always initiated from the recipient, or target, of assimilatory rules, and this fact leads to the elimination of iterative rule application.

d. Both SEARCH and COPY may have arbitrarily specified conditions on their application, and these formal distinctions allow for the analysis of empirical differences.

ISSN 1450–3417

We proceed shortly to the illustration of proposals (1a-d), and show how they can derive a wide variety of phenomena that are traditionally grouped under the rubric 'vowel harmony'. First we shall say a quick word about phonological rules and representations.

2. On Proceduralism and Precedence

In this contribution we propose a strongly procedural model for vowel harmony, and we expect to extend this analysis to other types of processes. We have provided arguments elsewhere (e.g., Reiss, in press) against non-procedural, constraint-based models. Our approach may appear to be at odds with the phonological *Zeitgeist*, given the popularity of Optimality Theory (OT), but closer inspection reveals that, confronted with the intransigence of opaque derivations, much current work in OT has retreated from the anti-procedural, two-level models that were first proposed. Kiparsky's (2000) stratal LPM-OT, which involves multiple levels in the generation of a surface form, and McCarthy's (2000) Harmonic Serialism, which is explicitly serialist and moreover involves iterative constraint application, are two examples of the return to a model that mimics the derivations of pre–OT work. The basic idea of a procedural approach is that the grammar should specify, not what is well-formed or ill-formed, but how to map an input to an output, as an explicitly characterized function. Our approach, maintains a procedural view, but eschews constraints altogether. Interestingly, our interpretation of phonological rules appears to simplify derivations in requiring fewer rule applications and fewer intermediate levels of representation than traditional models that required iterative rule application, a mechanism we avoid.

We accept Raimy's (2000) arguments and adopt his proposal that phonological strings are ordered sets of timing slots associated with feature bundles, and moreover that any ordering on features is induced from this order.[1] Formally, then, we take a phonological string to be a total order $\Sigma = \langle X, \preceq \rangle$, and the expression '$a \preceq b$' is read "the timing slot to which feature bundle *a* is associated *precedes* the timing slot to which feature bundle *b* is associated" — for short, "segment *a* precedes segment *b*."[2]

Following standard mathematical practice, we define immediate precedence as a special sub-case of precedence: $a \prec b \Leftrightarrow a \preceq b \ \&\ \forall\, c \neq a, c \preceq b \Rightarrow c \preceq a$. In words, *a* immediately precedes *b* if and only if *a* precedes *b* and for all *c* other than *a*, if *c* precedes *b*, then *c* precedes *a*. Reducing immediate precedence to a sub-case of precedence allows us to take the perspective that rules involving segmental adjacency are really just special cases of long-distance interactions. It

---

1 In fact, Goldsmith (1979: 28) explicitly states that tiers are ordered as well: "Each autosegmental level is a totally ordered sequence of elements, a_j^i: this is the j^{th} element on the i^{th} level. Call the set of segments on the i^{th} level L^i." Empirical considerations about contour tones, for example, support this view. Since this will not matter for our discussion, we do not address it further.

2 The totality of the order means that we do not allow precedence "loops" unlike in Raimy's initial account of reduplication.

then follows that long-distance interactions cease to have any special status since the machinery needed to formulate them is needed for adjacency as well. This view of long-distance dependency does go against the grain of most phonological research, where phonological rules are assumed to apply under adjacency, and non-local effects are either explained away, or else require special theoretical machinery.

3. The SEARCH Algorithm

We propose that assimilatory rules (and perhaps others) make use of a search procedure that stipulates a direction of search δ ('LEFT' or 'RIGHT') within a phonological string Σ, as well as initiation and termination criteria, denoted ς and γ, respectively. The algorithm, which is reminiscent of Chomsky & Halle's (1968: 344) approach to multiple rule application, is given in Figure 1.

SEARCH (Σ, ς, γ, δ):
1. Find all x in Σ subsumed by ς and index them: $\varsigma_0, \varsigma_1, \ldots, \varsigma_n$.
2. For each $i \in \{0, \ldots, n\}$:
   a. Proceed from ς_i through Σ in the direction δ until an element subsumed by γ is found.
   b. Label this element γ_i.
3. Return all coindexed pairs, (ς_i, γ_i).

Figure 1: The SEARCH algorithm

Thus an application of SEARCH will find one terminating point — the closest one in the appropriate direction — for each ς_i. The crucial point, however, is that SEARCH proceeding from two distinct starting points, ς_i and ς_j, may terminate on a common goal, returning pairs (ς_i, γ_i) and (ς_j, γ_j), where $\varsigma_i \neq \varsigma_j$, but $\gamma_i = \gamma_j$. Such a goal will bear multiple indices: $\gamma_{i,j}$.[3]

This property of SEARCH, in which multiple initiation points may come to be associated with a single goal segment, effectively eliminates the need for iterative application of harmony rules that spread a feature value in "local" steps. In one fell swoop, each harmonizing segment finds the closest instance of the relevant feature. As a simple illustration of this, consider the following abstract string, where x_1 and x_2 are of type X, and y_1 and y_2 are of type Y:

(2) $\Sigma = [x_1 \prec x_2 \prec y_1 \prec y_2]$

[3] Note that ς and γ are being used to refer to both types and tokens. Unindexed ς and γ are always feature specifications that define the type (natural class) of the initiating and terminating segments of the SEARCH algorithm, while indexed ς_i and γ_j are token segments subsumed by ς and γ.

Assume now that we have invoked the procedure SEARCH(Σ, X, Y, 'RIGHT'), or in words, "identify segments of type X and search to the right for segments of type Y." SEARCH will return the following set of pairs: $\{(x_{1i}, y_{1i,j}), (x_{2j}, y_{1i,j})\}$. That is, y_1 is the first element of type Y to the right of x_1, and it is also the first element of type Y to the right of x_2. The example makes it clear that the locality relations defined by SEARCH are non-reversible: Although y_1 is the closest element of type Y to the right of x_1, it is not the case that x_1 is the closest element of type X to the left of y_1. In fact, it is x_2 that is the closest element of type X to the left of y_1.

Traditionally, locality has been taken to be a symmetric relation, and one could simply say that a pair of segments *a* and *b* were in the relation of locality. Given the stance that we have adopted, in which locality is not a grammatical primitive, but is instead derived from a typed and directionally-specified SEARCH procedure, we can see that the traditional, simplistic view is no longer sufficient. Although it seems that we have complicated the phonology, this is in fact not the case, since locality has now been taken out of the grammar. Moreover, we shall see that the derived relation allows for a unified analysis of seemingly disparate and complex phenomena. This capturing of empirical generalizations is the true litmus test of the suitability of our modifications to the theory of phonological computation.

This property of searching linearly, but for objects of a particular type, is crucial to the existence of long-distance interactions in phonology. In fact, rather than viewing this discussion as an analysis of locality, it may be more useful to view it as an exploration of the mechanisms that allow for long-distance interactions in phonology.

Long-distance dependencies have been considered one of the defining features of human language, at least since *Syntactic Structures* (Chomsky 1957), and so we find it curious that so much ink has been spilled attempting to explain away such relations in phonology. Given our view of adjacency as a special case of long-distance dependency, and given the obvious parallels with syntax, we find no motivation for the eliminative reduction of long-distance effects in phonology.

4. SEARCH and COPY: Standard as Target

In the subsequent discussion we focus on feature-filling vowel harmony processes. Such processes involve filling in a feature value, [αF] onto a recipient vowel by copying [αF] from a donor segment elsewhere in the phonological string. As mentioned in proposal (1c), we claim that the recipient segment in the COPY operation (see Figure 2 below) is always the initiation point of the SEARCH algorithm. This discovery is made possible by recognizing the inherently asymmetric nature of the relation established by SEARCH.[4]

[4] A useful parallel is found in the Agree relation in syntax (e.g., Chomsky 2000b), in which a node with unvalued features probes for a matching feature. The mechanism motivating the filling-in of features in this view of syntax is Full Interpretation (cf. Chomsky 1995), but we do not intend this to be an explicit or implicit endorsement of a similar claim for phonology. See Nevins 2004 for an explicit attempt to link vowel harmony with syntactic Agree.

The proposal is stated in (3):

(3) *Big data claim*

Feature-filling vowel harmony involves *recipient* segments searching for and copying features from *donors*; donors do not search for and spread features to recipients.

Whether this claim is valid in other assimilatory processes is a question for further research — in this article we will limit our empirical domain and show that (3) appears to be valid. We will also suggest that (3) can be explained; in other words, it is not an arbitrary fact about phonological computation.

COPY (ς_i, γ_i, αF, C): Identify αF on γ_i and assign αF to ς_i if the set of conditions C on γ_i are satisfied.

Figure 2: The COPY operation

5. Basic Illustration: Turkish

In this section, we show how the SEARCH and COPY approach generates the well-known basic[5] vowel harmony patterns of Turkish. Turkish has suffixes whose vowels alternate between [e] and [a] in agreement with the value of [BACK] on the preceding vowel. An example is the plural marker *–ler/–lar*. We assume that the vowel in this suffix is underlyingly [–HIGH, –ROUND] and that the value for [BACK] is filled in by applying SEARCH and COPY as outlined above. We use V here to denote a vowel that is unspecified for the termination criterion, and follow this convention unless otherwise specified.[6] The direction of SEARCH is leftward, and γ is [αBACK], that is, any token of a value for the [BACK] feature.

(4) *SEARCH in Turkish [BACK] harmony*

a. ς = V
b. δ = 'LEFT'
c. γ = [αBACK] (a vowel with any [BACK] specification)

COPY then assigns the value [αBACK] that is found to the suffix vowel. This analysis generates the plural forms seen in the first column of Table 1.

---

5 Note that the facts of Turkish harmony are more complex than we show here. We will return to the issue when we discuss consonant-vowel interactions in vowel harmony, discussing data highlighted by Nevins (2004).

6 See Reiss 2003 for the mechanism by which it is possible to refer to a segment that necessarily lacks a value for a given feature.

	–NOM.PL	*–GEN.SG*	*–PL–GEN*	*gloss*
a.	ip–ler	ip–in	ip–ler–in	'rope'
b.	kɨz–lar	kɨz–ɨn	kɨz–lar–ɨn	'girl'
c.	sap–lar	sap–ɨn	sap–lar–ɨn	'stalk'
d.	yüz–ler	yüz–ün	yüz–ler–in	'face'
e.	son–lar	son–un	son–lar–ɨn	'end'

Table 1: Turkish vowel harmony data

Note that the possessive suffix, like other Turkish suffixes with high vowels, shows a four-way alternation. However, we can assume that the same process that accounts for the values of back for the non-high suffixes accounts for it in the suffixes with high vowels. The representation of the vowel in the possessive suffix is [+HIGH], and since it has no value for [BACK] these high vowels also serve as initiation points for the SEARCH and COPY operations formulated above.

However, the high vowels also agree with the preceding vowel with respect to the feature [ROUND]. We can derive this distribution by again applying a SEARCH and COPY. In this case SEARCH is initiated by a vowel without a specification for [ROUND]:

(5) *SEARCH in Turkish [ROUND] harmony*

a. ς = V
b. δ = 'LEFT'
c. γ = [αROUND]

COPY will then assign the value of [ROUND] found on γ_i to ς_i. Note that this process makes no reference to the fact that the targets of [ROUND] harmony are all [+HIGH] — this follows from the fact that the [+HIGH] suffixes are lexically unspecified for [ROUND], whereas the [–HIGH] suffixes are lexically [–ROUND] and thus do not serve as initiation points for this invocation of SEARCH.

We now turn to the forms in the third column, the possessive plurals, which show both of the suffixes we have just considered. Applying the SEARCH and COPY rules for both [BACK] and [ROUND] produces exactly the desired result. In this case we have no evidence for which process applies first, and we will illustrate applying the process for [BACK] before the process for [ROUND]. The important details of the example are illustrated in Figure 3.

(ROOT)	≺	l	≺	V	≺	r	≺	V	≺	n
				–HIGH –ROUND				+HIGH		

Figure 3: Schematic representation of unharmonized Turkish root-suffix combination

Following the root are the plural marker and possessive suffix, respectively. The plural affix contains a vowel specified only for height and rounding, while the possessive has a vowel specified only for height.

The vowels of *both* suffixes are starting points for the [BACK] harmony process, since neither is specified for [BACK]. SEARCH starts at each of these standards and finds the left-closest segment that is specified for backness, whether it is [+BACK] or [–BACK]. Assuming the root vowels are all specified for [BACK], both suffixes will have SEARCH terminate on the final root vowel and will copy its specification for [BACK] — both recipients make use of the same donor. There is no need to apply the rule iteratively, to first give the plural suffix a value for [BACK] and then copy the value from the plural to the possessive suffix.[7]

We see that the terminating segment is "local" in the sense that it is the first eligible donor found by SEARCH. We propose that this is the only sense of locality that is relevant to phonological computation.

Now consider rounding harmony in the possessive plural. The vowel of the plural is already specified [–ROUND] so it cannot serve as an initiating point for SEARCH, thus it cannot be targeted by the rule. The vowel of the possessive, on the other hand, is not specified for roundness. SEARCH is initiated at that vowel and looks for the first specification of [ROUND] to the left. It always terminates on the [–ROUND] vowel of the plural marker, and so we only find [–ROUND] versions of the possessive suffix when it follows the plural suffix.

The two feature-filling rules of Turkish vowel harmony are thus as follows:

(6) *Turkish vowel harmony*

a. [BACK]
   i. From [ς : V] SEARCH left for [γ : αBACK].
   ii. COPY [αBACK] to ς.

b. [ROUND]
   i. From [ς : V] SEARCH left for [γ : αROUND].
   ii. COPY [αROUND] to ς.

There is no evidence for ordering between the rules. The rules do not apply iteratively to their own outputs, since each SEARCH can occur simultaneously to find the first source to copy from towards the left.

6. Accounting for Neutral Vowels

Scholars of vowel harmony have long struggled with the phenomena of *opaqueness* and *transparency*. In the case of opaqueness, a non-alternating "neutral" vowel blocks the spread of [αF] and spreads its own feature value. Transparent neutral vowels, on the other hand, appear to be invisible to the

7 A reviewer remarks that our simultaneous application of harmony to all suffixes implies that we are adopting a "phonology after morphology" approach to the grammar. These kinds of examples are convenient for the illustration of simultaneous application — polysyllabic suffixes would do as well — and aren't meant as an explicit or implicit endorsement of any higher-level architectural design choices.

harmonic process, allowing features to spread "through" them. Generally, the explanations for neutral vowels invoke either (i) a special property inherent to the vowels themselves or (ii) additional rules or constraints that apply only to these vowels (Bakovic & Wilson 2000: 45). It is interesting also to note that there are so few unified accounts of both types of neutrality, in fact the properties of opaque vs. transparent vowels — or the rules that apply to them — are often claimed to differ in important ways. We shall show below how our theory achieves this unification elegantly, without appeal to "special" properties of neutral vowels or positing unmotivated theoretical machinery.

We assume that in a language with both harmonizing and neutral vowels, those vowels that alternate have no value specified underlyingly for the harmonic feature [F] and surface as [+F] or [–F] depending on the specification of the vowels with which they harmonize, while non-alternating vowels fail to undergo harmony because they are *underlyingly already specified* for the harmonic feature, and the relevant rule is feature-filling. We see, then, that there is nothing special about neutral vowels. In fact, they could be considered the more "normal" vowels, being underlyingly more specified than their alternating counterparts.

Since we treat all non-alternating vowels as being underlyingly fully specified, it is clear that the terms OPAQUE and TRANSPARENT are stripped of any theoretical significance. Ultimately, we will show that these labels reflect differences in properties of *rules*, rather than intrinsic properties of the vowels themselves.

In the following sections, we show how our model accounts for neutral vowels without recourse to a difference between consonant and vowel place features or nodes (cf. Clements & Hume 1995), or other enriched representational apparatus. Both opaqueness and transparency can be shown to follow from the nature of the rules applied to the vowel representations we posit.

7. Opaqueness

We propose that the situation we observed in Turkish, in which the [–ROUND] value on the plural suffix "blocks" access to the value for [ROUND] on the preceding root vowel, sheds light on the phenomena that characterize opaque vowels in harmony systems.

A traditional interpretation of the Turkish phenomena might say that the vowel of the plural is opaque in the sense that it prevents the harmonic feature of the root vowel from spreading across it to the high suffix vowels. Such an account typically appeals to special representational properties of the vowel in question, or, in an autosegmental framework, to a ban on crossing association lines.[8] Our own derivation of the opaque behaviour did not require either of these theoretical devices, and in fact only appealed to independently-motivated properties of phonological rules and representations: segmental underspecifi-

---

8 See Coleman & Local 1991 for an argument that the NO CROSSING CONSTRAINT is incoherent.

cation and linear string scanning.[9] The vowel of the plural just happens to already have a value for [ROUND], and this value is visible to the SEARCH procedure which scans for any value [αROUND]. We also did not appeal to the inventory of vowels in the language — there are round, non-high vowels in Turkish, but they happen not to participate in the harmonic alternations. In other words, opaqueness among vowels is not dependent on the structure of the surface vowel inventory, since the Turkish non-high vowels /a, e/ have round counterparts /o, ö/ — they are harmonically paired, to use a current phrase (cf. Bakovic 2003 and Mahanta 2005, *inter alia*) — and yet they are opaque. We think that this general approach can be applied unchanged to cases that are viewed as more typical examples of opaqueness in harmony systems, such as the [ATR]-opaque low vowel in Tangale.

Tangale is a Chadic language with tongue root harmony. The /a/ vowel fails to harmonize, and it furthermore blocks copying of a harmonic feature to its left.

	underlying	*surface*	*gloss*
a.	/seb-U/	[sebu]	'look' (IMP)
b.	/kɛn-U/	[kɛnʊ]	'enter' (IMP)
c.	/peer-na/	[peerna]	'compelled'
d.	/pɛd-na/	[pɛdna]	'untied'
e.	/ɗob-Um-gU/	[ɗobumgu]	'called us'
f.	/ɗib-na-m-gU/	[ɗibnamgʊ]	'called you' (PL)

Table 2: Tangale [ATR] harmony

(van der Hulst & van de Weijer 1995)[10]

In Table 2, items (a) and (b) show that values of the feature [ATR] spread rightwards (or are copied from the left, on the present account), while (c) and (d) show that /a/ fails to alternate. Item (f) is the crucial piece of data, showing that /a/ not only fails to alternate, but in fact behaves as a copying source for its own [–ATR] value, blocking the copying of [+ATR] from preceding vowels. These data are all accounted for straightforwardly by assuming that the vowels denoted by /U/ are underspecified with respect to the feature [ATR], and that such vowels serve as the initiating points for SEARCH in the [ATR] harmony rule. SEARCH terminates on any vowels specified for [ATR] and this includes /a/, the only low vowel in the language, which happens to be [–ATR].

(7) *Tangale [ATR] harmony*

a. From [ς : V] SEARCH left for [γ : αATR].
b. COPY [αATR] to ς.

9 See Inkelas 2006 on underspecification. A linear scanning procedure is at least necessary for identifying potential environments of application for phonological rules, or constraint-violation locations in OT.

10 Item (e) is from Bakovic (2003).

Because /a/ is underlyingly specified [–ATR], it cannot initiate the SEARCH for this feature-filling rule. Also because it is underlyingly specified [–ATR], it terminates SEARCH initiated by vowels to its right. This provides a complete explanation for the opaqueness of this vowel: Opaqueness follows from the nature of the rule system and the straightforward representation of the vowels.

To reiterate our point about inventories, Turkish [e] and [a] are "paired" with round counterparts /ö/ and /o/, respectively. However, this surface fact is irrelevant to the behavior of these vowels with respect to the [BACK] and [ROUND] harmony systems. In alternating suffixes, these surface vowels reflect a feature bundle which is not specified for [BACK], and thus does not cause a SEARCH for [αBACK] to terminate; but it is specified for [–ROUND], so it does cause a SEARCH for [αROUND] to terminate, giving rise to 'opaqueness'. The Tangale /a/ is specified [–ATR] and thus causes a SEARCH for [αATR] to terminate, giving rise to opaqueness.

The Tangale opaque vowel, like the other harmony triggers, is fully specified, and there are no [ATR] underspecified vowels of the same height as the opaque one. The parallel situation holds in Turkish — there are no [–HIGH] vowels in Turkish underspecified for [ROUND].

8. COPY is Independent of SEARCH

Thus far we have implicitly assumed that the description of the termination criterion for SEARCH and the description of what COPY copies are identical. For example, in Turkish [BACK] harmony, SEARCH looks for a specification for [BACK] and copies it onto the standard. In Tangale, SEARCH looks for an [ATR] value which COPY copies. However, there is no *a priori* reason to assume that the specification of γ and the description of what is to be copied must be identical.[11] In other words, we are justified in breaking down harmony into these two separate processes.

This mismatch between the specification of what is targeted by SEARCH and what is targeted by COPY gives rise in languages like Finnish and Wolof to what is called *transparency*, a kind of neutrality different than the opaqueness seen in Tangale, which we discuss in the following section.

9. Transparency

So called transparent vowels also fail to alternate, and thus by our previous assumptions — but contrary to most of the literature — must underlyingly have a value specified for the harmonic feature. In contrast to opaque vowels, however, they appear to allow harmonic features to be copied across them. In the following subsections we will show with data from several languages that the

[11] It does, however, seem to be the case that γ must be subsumed by the specification of the copied value. Otherwise the possibility exists of attempting to copy a feature value that is not present.

typical view of transparency as a unified phonological phenomenon is erroneous, and that transparent behaviour in vowel harmony has at least three distinct sources:

(8) *Sources of transparent vowel behavior*

a. Conditions on the target of SEARCH
b. Conditions on the target of COPY
c. Rule ordering

9.1. Transparency in Wolof via Conditions on SEARCH

In the Wolof system of [ATR] harmony the two high vowels /i, u/ are transparent to the harmony process, as the data in Table 3 show.[12]

	underlying	*surface*	*gloss*
a.	/toxi-lEEn/	[toxileen]	'go & smoke' (IMP)
b.	/seen-uw-OOn/	[seenuwoon]	'tried to spot'
c.	/tɛkki-lEEn/	[tɛkkilɛɛn]	'untie' (IMP)
d.	/tɛɛr-uw-OOn/	[tɛɛruwɔɔn]	'welcomed'

Table 3: Wolof [ATR] harmony

As usual, we assume that /i/ and /u/ are underlyingly specified for [ATR], as they do not alternate. However, the suffixes that follow these vowels appear to copy their [ATR] specification from the vowel before the /i/ or /u/. Why don't these vowels terminate a leftward SEARCH for [ATR] initiated by a vowel to their right? The answer we propose is simply that conditions on SEARCH (i.e. the initiating and terminating criteria) need not be singleton features, but are stated in terms of natural classes, that is, potentially complex conjunctions of phonological features. In the case of Wolof, SEARCH has as terminating criterion [αATR, –HIGH], that is, SEARCH will only terminate at a non-high vowel that is specified for [ATR].

(9) *Wolof [ATR] harmony*

a. From [ς : V] SEARCH left for [γ : –HIGH, αATR].
b. COPY [αATR] to ς.

This kind of featural specification is widespread in phonological processes — one language may have a rule affecting all vowels in a particular environment, whereas another language affects only [+HIGH] vowels in the same environment. As another example, a language may have a rule affecting *voiced obstruents*, a

12 The data are from Archangeli & Pulleyblank (1994), but we have standardized the transcription. Small capital letters denote vowels that do not alternate, and hence have no [ATR] specification. The symbols [i, u] denote high vowels that are [+ATR].

description that must be specified with a conjunction of features.

As in the case of opaqueness, we do not require any new representational machinery to capture transparency. In fact, there is nothing special about transparent vowels at all in Wolof; their transparency is not a property of the vowels themselves, but rather it follows from the conjunctive feature specification of γ in the [ATR] harmony rule.

9.2. *Transparency in Hungarian via Rule Ordering*

We addressed above the question of whether the presence of opaque vowels depends on the vowel inventory of a particular language. Transparency has also been said to depend on the lack of a harmonic pair. Basically vowels are said to be neutral in one of these two ways if they are not "matched" in the vowel inventory with respect to the relevant harmonic feature. Thus, much work on harmony makes crucial reference to the vowel inventory of a language in formulating a computational analysis of its harmony patterns. While inventories are commonly referred to by phonologists in describing linguistic forms, we believe that they play no explanatory role as part of the (mental) grammar of a language. After all, the inventory, if it refers to underlying vowels, is just a redundant catalog of the contents of the lexicon.

Most harmony work that refers to inventories directly or by referring to contrastive features does so because neutral vowels *tend* to be ones that do not have a harmonic mate in the surface inventory. We believe that neutrality cannot actually be explained by reference to the surface inventory, since this inventory is derived by the phonology — it can't be the case that the derived inventory determines the phonology.

To support the idea that the phonology of a language demands an analysis, not just a superficial catalog of surface segments, we present some observations concerning the surface vowel [e] in Hungarian. All Hungarian data are from Siptár & Törkenczy (2000).

The non-low, front unrounded vowels in Hungarian can be transparent to vowel harmony. These are orthographic *i, í, e, é*. An example is found in the deverbal adjective forming suffix *–ékony/–ékeny: gyúlékony* 'flammable', *közlékeny* 'talkative'. The first suffix vowel *é* is transparent, whereas the second vowel harmonizes for the feature [BACK] — it is [o] when the last root vowel is [+BACK] /ú/, and [e] when the last root vowel is [–BACK] /ö/.[13]

The features of the transparent vowels are shown in Table (4).

---

[13] We have not found any explicit discussion of the failure of rounding harmony in *közlékeny*. It appears that the *o/e* alternation in this suffix, which does not include a front rounded [ö] version, requires an underlying representation different from any of the other alternating vowels — [e] alternates with either [ɔ] (orthographic *a*) or with [o] and [ö].

A full analysis is not possible here, but it looks like it might work to posit underlying [–HIGH, –LOW, –ROUND] as the representation of this vowel. The feature-filling rule posited below would fill in values for [BACK]; another feature-filling rule would provide [ATR] in the context [BACK]; and then a feature-changing rule would change [–ROUND] to [+ROUND] on a non-low, [ATR] vowel which is [BACK].

orthography	*IPA*	*features*	*length*
i	[i]	[+HI, –LO, –BK, –RD, +ATR]	short
í	[i:]	[+HI, –LO, –BK, –RD, +ATR]	long
e	[ɛ]	[–HI, –LO, –BK, –RD, –ATR]	short
é	[e:]	[–HI, –LO, –BK, –RD, +ATR]	long

Table 4: Transparent vowels of Hungarian

Note that, unlike the transparent vowels of Finnish, some of these surface vowels can also be the surface manifestation of alternating vowels. Short *e* surfaces in alternation with *a*, as in the inessive suffix: *dobban* 'in a drum', *szemben* 'in an eye'. It also surfaces in alternation with the tense round mid vowels *o/ö*. This pattern is seen in the superessive suffix *–en/–on/–ön*: *szemen* 'on an eye', *tökön* 'on a pumpkin', *dobon* 'on a drum'. The long *é* surfaces in alternation with *á*, as in the translative suffix *–vá/–vé* (the *v* assimilates to a preceding consonant): *dobbá* '(turn) into a drum', *szemmé* '(turn) into an eye, *tökké* '(turn) into a pumpkin'.

There is no problem with the fact that a surface vowel such as *é* can correspond to both a vowel of a harmonizing suffix and a non-alternating "transparent" vowel. The non-transparent cases of alternating *i, í, e, é* just represent surface realization of vowels that are partially underspecified underlyingly. These missing values are filled in by rule.

In the latter, "transparent" case, the vowel is fully specified (with the values in Table 4), and it does not alternate. Transparency effects in these non-alternating vowels can be derived using the simple mechanism of rule-ordering.

(10) *Hungarian [BACK] harmony, version 1*

a. i. From [ς : V] SEARCH left for [γ : +BACK]
   ii. COPY [+BACK] to ς

b. i. From [ς : V] SEARCH left for [γ : -BACK]
   ii. COPY [–BACK] to ς

Rule (10a) copies [+BACK] from the first [+BACK] found to the left. If one is found, then there will be no vowels left underspecified for [BACK] and thus feature-filling rule (10b) cannot apply. If no [+BACK] is found, then unspecified vowels will still initiate the SEARCH of rule (10b) and they will always find [–BACK], which can be copied. The innovation of this proposal is that the so-called 'transparent' vowel is only transparent due to the condition on (10a), but in fact, the apparent transparent vowel is the termination vowel of the SEARCH of (10b). Transparency again is epiphenomenal — a result of a particular system of rules and representations.

To reiterate, the surface vowel *é*, for example, corresponds to both a non-harmonizing, underlyingly fully specified transparent vowel, and to one surface manifestation of a harmonizing, underlyingly partially underspecified vowel. Obviously, the dual behavior of these "transparent" vowels can have nothing to do with the surface vowel inventory, since distinct underlying vowels are merged on the surface.

However, this account of Hungarian is still incomplete.[14] Consider first the disharmonic stems in Table 5, in which the final vowel is front, but a member of the "transparent" class.

	form	*gloss*
a.	papír–nak	'paper'–DAT
b.	kábít–om	'daze'–1SG.DEF
c.	gumi–nak	'rubber'–DAT
d.	Tomi–nak	'Tom'.DIM–DAT
e.	kávé–nak	'coffee'–DAT
f.	bódé–tól	'hut'–ABLAT

Table 5: Disharmonic stems with transparent vowels

Since our rules are ordered to first seek [+BACK], the transparent vowels will be skipped and the preceding [+BACK] vowels will terminate the search, and [+BACK] will be copied to the initiator of SEARCH, i.e. the suffix vowel.

Conversely, the disharmonic stems in Table 6 contain, in (a) and (b), final front round vowels, which are not transparent.

	form	*gloss*
a.	sofőr–nek	'driver'–DAT
b.	parfüm–nek	'perfume'–DAT
c.	büró–nak	'bureau'–DAT
d.	béka–nak	'frog'–DAT

Table 6: Disharmonic stems with opaque vowels

Note that the suffixes agree with the immediately preceding vowel even in (a) and (b) where that vowel is [–BACK], and not the [+BACK] vowel that precedes. Thus, the ordered rules in (11) will generate the wrong output for these forms, although they will work for (c) and (d). However, exploiting the theoretical machinery that we already have in place for cases like that of Wolof, we can see that making SEARCH terminate on more narrowly specified segments will solve this problem. The following set of harmony rules correctly generate all of the forms that we have considered:

(11) *Hungarian [BACK] harmony, final version*

a. i. From [ς : V] SEARCH left for [γ : +ROUND, αBACK].
   ii. COPY [αBACK] to ς.

b. i. From [ς : V] SEARCH left for [γ : αBACK].
   ii. COPY [αBACK] to ς.

[14] The data in the next two examples are well known, but we acknowledge Benus (2005) as the immediate source.

The fact that two different rules can potentially fill in the value of [BACK], may appear to be inelegant, but the mechanisms used are independently necessary: Rule ordering is a basic feature of the derivational model we assume, and an account that assumes underspecification for transparent vowels needs additional rules to fill in their surface values by the end of the derivation. Moreover, our account provides a principled explanation for the fact that "transparent" vowels trigger [–BACK] harmony when there are no other vowel types in the word.[15]

9.3. *Taking Stock*

We have seen thus far two of the possible sources of "transparent" behaviour in vowel harmony: In Wolof the featural specification of γ introduces the possibility of long-distance termination of SEARCH, while in Hungarian the ordering of two feature-filling rules with different termination criteria but identical targets of COPY leads to a "transparent" value being searched for and copied after a "non-transparent" value. We turn now to the final source of transparency, which takes the form of conditions on the application of the COPY operation.

9.4. *Kirghiz: Conditions on COPY*

Kirghiz, another Turkic language, displays a quirky exception to its otherwise general pattern of palatal and labial harmony: Non-high vowels do not assimilate in rounding to high back round vowels, but do assimilate to high *front* round vowels. This is shown in Table 7:

	accusative	*dative*	*gloss*
a.	taʃ–tɨ	taʃ–ka	'stone'
b.	iʃ–ti	iʃ–ke	'job'
c.	utʃ–tu	utʃ–ka	'tip'
d.	konok–tu	konok–ko	'guest'
e.	köz–tü	köz–gö	'eye'
f.	üy–tü	üy–gö	'house'

Table 7: Kirghiz vowel harmony data

The crucial data in Table 7 are the dative forms in (c) and (d), in which the *–kV* suffix does not copy [+ROUND] from a preceding /u/, but does copy it from a preceding /o/.

Since all alternating vowels assimilate in backness to the preceding vowel, a simple rule is sufficient, as in (12a). In order to deal with the failure of /u/ to trigger round harmony in a non-high vowel, we need two separate rules like (12b-c), which will assign [+ROUND] to a [–HIGH] vowel when the preceding vowel is [–HIGH] or when it is [–BACK], respectively.

[15] Ignoring, of course, the exceptional stems in Hungarian that take [+BACK] harmony despite having only transparent vowels, e.g., *híd–nak* 'bridge'–DAT.

(12) *Kirghiz rules*

a. i. From [ς : V] SEARCH left for [γ : αBACK].
ii. COPY [αBACK] to ς.
b. i. From [ς : V] SEARCH left for [γ : αROUND].
ii. COPY [αROUND] to ς *if* γ is [–HIGH].
c. i. From [ς : V] SEARCH left for [γ : αROUND].
ii. COPY [αROUND] to ς *if* γ is [–BACK].

These rules generate the observed patterns of alternation. Since SEARCH looks for the immediately preceding vowel, there is no chance of observing transparency or opaqueness effects. However, this pattern is similar to transparency in that a given vowel, /u/, which we assume is specified for a particular value, cannot transmit that value to an underspecified vowel that probes it. In contrast to Wolof, where the mechanism used to generate the inertness of [+ATR] on high vowels was to put conditions on SEARCH, in Kirghiz we put conditions on the application of the COPY operation: COPY only applies if the segment that terminates SEARCH meets the conditions in (12b) or (12c). Note, moreover that if (12b) applies, then (12c) cannot, as ς will no longer be underspecified with respect to rounding. The ordering of (12b) and (12c) cannot be determined.

10. Understanding Conditions on SEARCH and COPY

The difference between imposing conditions on the target of SEARCH versus the target of COPY is perhaps non-obvious, and so we take a moment here to discuss it further. Consider the following abstracted versions of the relevant harmony processes:

(13) *Schemata for conditional harmony*

a. SEARCH left for [*x*,*y*]; COPY *x*. (e.g., Wolof)
b. SEARCH left for [*x*]; COPY *x* if [*x*,*y*]. (e.g., Kirghiz)

The difference between templates (13a) and (13b) is illustrated by the following scenarios. Suppose you are told to go out into the world, find a man with a hat, and take his hat. On the assumption that there are such things as men with hats and that they are findable, you will always return with a hat. But the outcome is potentially different if you are told to go out, find a *person* with a hat, and take the hat *only if that person is a man.* You may in this case return hatless, if the first behatted person you met was a woman. The first task involved a condition on the search termination — take the hat of the first person you meet who is both a man *and* a hat-wearer; the second involved a condition on the hat-taking (COPY) operation — take the hat of the first hatwearer, only if that person is a man.

As it turns out, our account of harmony has thus far glossed over a potentially important point by unintentionally conflating the (21a) template with SEARCH terminated by a feature singleton. A rule like “From [ς : V] SEARCH left for [γ : –HIGH, αATR]” in Wolof (cf. 13a) fails to specify that it is a *vowel* that is being searched for. That is, the specification for γ should in fact read [+VOCALIC, –HIGH, αATR].[16] On the assumption that vowels and consonants can share at least some features, this tacit omission of vowel specification can lead to incorrect predictions about output forms. We will see such a case in the following section.

11. Consonant–Vowel Interactions: Turkish Laterals

In this section we briefly examine a less-studied aspect of harmony, the interaction of consonants and vowels, exemplified with laterals in Turkish. We will show that the theoretical machinery we already have in place allows us to account straightforwardly for well-known cases, provided we are sufficiently explicit in specifying our rules and representations.

Turkish has both palatalized and non-palatalized laterals, and these have been shown to interact with the general pattern of backness harmony.

	bare	*inflected*	*gloss*
a.	usul^y	usul^y–ü	‘system’–ACC.SG
b.	petrol^y	petrol^y–ü	‘petrol’–ACC.SG
c.	sual^y	sual^y–i	‘question’–ACC.SG
d.	okul	okul–u	‘school’–ACC.SG
e.	karakol	karakol–u	‘police.station’–ACC.SG
f.	tʃatal	tʃatal–ɨ	‘fork’–ACC.SG
g.	petrol^y	petrol^y–de	‘petrol’–LOC.SG
h.	meʃgul^y	meʃgul^y–düm	‘busy’–PAST.1.SG

Table 8: Turkish palatalized and non-palatalized laterals

(Nevins 2004: 40)

Simply stated, alternating suffixes surface with front harmony if the final consonant is a palatalized lateral, even if the preceding vowel is [BACK]. If we look back at the rule templates in the previous section, and keep in mind that feature-matching in phonology is done by subsumption, the appropriate rules for the Turkish case are clear:

(14) *Turkish harmony, final version*

a. [BACK]

i. SEARCH left for [αBACK].

ii. COPY [αBACK].

[16] We assume that the vowel/consonant distinction is featurally specified. The exact mechanism does not matter for our purposes.

b. [ROUND]

i. SEARCH left for [+VOCALIC, αROUND].

ii. COPY [αROUND].

The first rule looks for *any* instance of [BACK] on any segment (i.e. vowel or consonant), whereas the second rule looks for [ROUND] exclusively on vowels. Thus, our initial formulation of Turkish back harmony in (6a) was correct in its form, but essentially by accident. The absence of a [VOCALIC] specification in part (14ai) above is crucial to a proper understanding of the role that consonants play in Turkish harmony.

Note that the account given above generalizes straightforwardly to all cases of consonant-vowel interaction in assimilatory processes. This eliminates the need for use of consonant and vowel features that are sometimes the same and sometimes not (cf. Spencer 1996).

12. Discussion: Phonology as Grammar

The present article is in many ways non-standard. It is not written from the perspective of Optimality Theory, the dominant theoretical approach to generative phonology for about a decade, and yet the framework on which it is constructed eschews many of the assumptions of so-called Classical Generative Phonology (viz. *SPE* and its descendants, up to and including Feature Geometric approaches). For these reasons we take a moment here to address some typical objections to our approach and summarize the motivations for the positions we have taken.

The most controversial aspect of this work for phonologists is likely to be our avoidance of any argument or analysis based on traditional typological and functionalist notions of markedness. Instead we adopt a "substance-free" approach, in which the computational system has no access to (and hence makes no use of) the phonetic substance of speech. The point is highlighted by Chomsky (2000a), who points out that it is a contingent fact that generative grammars give rise to language in humans, and that another creature may have a generative grammar that interfaces with completely different performance systems, for example, locomotion (cf. Lieberman 2000 on the "grooming grammars" of mice).

A puzzling type of comment evoked by work like this runs something like "You posit quantification, algebraic representations with variables, etc., and anything else you want, so you are just willing to posit any computational power at all in the phonology. If you do that, then phonology is not special in any way." In fact, the idea *is* to ascribe to the phonology any computational power it seems to need — but no more. Our claim is that a procedural approach to vowel harmony, and perhaps all assimilatory processes, *minimally requires* ordered representations and operations akin to (i.e. with at least as much computational power as) SEARCH and COPY. Not providing the phonology with the power it requires seems like a dead end if we are trying to understand what phonology is.

Another potential criticism of this contribution is that our examples merely demonstrate that we are clever enough to create a notational system that gets the results we are looking for. The alternative, not being able to be explicit about our claims and their consequences, seems unattractive at best. Moreover, we take a realist view of our notation — we develop notation that expresses what we assume to be the computational mechanisms used by the language faculty. As a recent example of how a simple decision to take notation seriously leads to theoretical insight, consider Raimy's (2000) explicit encoding of precedence relations in phonological representations. Either we can say that Raimy is "merely inventing a clever notation" or that he is making explicit the relations that the grammar has access to.

In general, the examples we have used to illustrate our approach are well-known and relatively simple, so our contribution offers little satisfaction for the reader looking for exotic data. This choice was a conscious one, since we adopt the view that the goal of particular sciences is to construct intelligible theories that yield insight into some narrowly circumscribed domain of the world of experience and observation. The data are typically too complex to be directly intelligible, and so it makes sense to start building our models with simple examples. Once the intelligibility and coherence of these models have been determined, we are in a position to move on to more complex phenomena. The notion that our data are too simple reduces to the suggestion that phonology has advanced enough that we no longer need to bother with such examples. We disagree.

We have aimed to provide a novel, yet simple account of phenomena that are fairly well-known by developing a rule-based framework with a minimum of ontological structure. The main contributions we hope to have made are:

(A) a novel, unified treatment of neutral vowels;

(B) clarification of the notion of locality in phonology;

(C) some insight into target/trigger relations in phonological processes;

(D) some ideas about the logical structure of rules.

In relation to (B), we remark here on the importance of distinguishing between descriptive and explanatory generalizations. Although putative "locality" effects are ubiquitous in phonology, we have showed that they are not properties of Universal Grammar *per se*, but rather are what Chomsky (2005) calls "third factor" effects, that is, they follow from extralinguistic facts about the nature of computation and search. This refinement of the boundary between ontological and epistemological facts is a clear sign of progress in the study of the properties of Universal Grammar.

References

Archangeli, Diana & Douglas Pulleyblank. 1994. *Grounded Phonology* (Current Studies in Linguistics 25). Cambridge, MA: MIT Press.

Bakovic, Eric. 2003. Vowel harmony and stem identity. *San Diego Linguistics Papers* 1, 1-42. [Also available online at the Rutgers Optimality Archive: http://roa.rutgers.edu/files/540-0902/540-0902-BAKOVIC-0-6.PDF.]

Bakovic, Eric & Colin Wilson. 2000. Transparency, strict locality, and targeted constraints. *West Coast Conference in Formal Linguistics (WCCFL)* 19, 43-56.

Benus, Stefan. 2005. Dynamics and transparency in vowel harmony. New York: New York University dissertation.

Chomsky, Noam. 1957. *Syntactic Structures* (Januar Linguarum, Series Minor 4). The Hague: Mouton.

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.

Chomsky, Noam. 2000a. *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

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

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

Chomsky, Noam & Morris Halle. 1968. *The Sound Pattern of English*. Cambridge, MA: MIT Press.

Clements, George N. & Elizabeth V. Hume. 1995. The internal organization of speech sounds. In Goldsmith (ed.), 245-306.

Coleman, John & John Local. 1991. The 'No Crossing Constraint' in autosegmental phonology. *Linguistics and Philosophy* 14(3), 295-338.

Goldsmith, John A. 1979. *Autosegmental Phonology*. New York: Garland Press.

Goldsmith, John A. (ed.) 1995. *The Handbook of Phonological Theory* (Blackwell Handbooks in Linguistics). Malden, MA: Blackwell.

van der Hulst, Harry & Jeroen van de Weijer. 1995. Vowel harmony. In Goldsmith (ed.), 495-534.

Inkelas, Sharon. 2006. Underspecification. In Keith Brown (ed.), *Encyclopedia of Language and Linguistics*, 224-226. Oxford: Elsevier.

Kiparsky, Paul. 2000. Opacity and cyclicity. *The Linguistic Review* 17(2–4), 351-367.

Lieberman, Philip. 2000. *Human Language and Our Reptilian Brain: The Subcortical Bases of Speech, Syntax, and Thought* (Perspectives in Cognitive Neuroscience). Cambridge, MA: Harvard University Press.

Mahanta, Shakuntala. 2005. Morpheme Realization and Directionality in Vowel Harmony. In Anna Asbury, Ivana Brasileiro & Shakuntala Mahanta (eds.), *UiL-OTS Yearbook 2005*. Utrecht: Utrecht Institute of Linguistics Publications, 51-64.

McCarthy, John. 2000. Harmonic serialism and parallelism. *North Eastern Linguistics Society (NELS)* 30, 501-524.

Nevins, Andrew. 2004. Conditions on (dis)harmony. Cambridge, MA: MIT dissertation.

Prince, Alan & Paul Smolensky. 2004. *Optimality Theory: Constraint Interaction in Generative Grammar*. Malden, MA: Blackwell.

Raimy, Eric. 2000. *The Phonology and Morphology of Reduplication* (Studies in Generative Grammar 52). Berlin: Mouton de Gruyter.

Reiss, Charles. 2003. Deriving the feature-filling/feature-changing contrast: An application to Hungarian vowel harmony. *Linguistic Inquiry* 34(2), 199-224.

Reiss, Charles. In press. Constraining acquisition without constraints. In Andrew Nevins & Bert Vaux (eds.), *Rules, Constraints, and Phonological Phenomena.* Oxford: Oxford University Press.

Siptár, Péter & Miklós Törkenczy. 2000. *The Phonology of Hungarian* (The Phonology of the World's Languages). Oxford: Oxford University Press.

Spencer, Andrew. 1996. *Phonology: Theory and Description* (Introducing Linguistics 1). Oxford: Blackwell.

Frédéric Mailhot
Carleton University
Institute of Cognitive Science
1125 Colonel By Drive
Room 2201, Dunton Tower
Ottawa
Ontario K1S 5B6
Canada
fmailhot@connect.carleton.ca

Charles Reiss
Concordia University
Classics, Modern Languages and Linguistics
1455 de Maisonneuve West
Room H-663
Montreal
Quebec H3G 1M8
Canada
reiss@alcor.concordia.ca

Categorical Acquisition: Parameter-Setting in Universal Grammar

Rosalind Thornton & Graciela Tesan

Two triggering models of parameter-setting, the Hierarchical Acquisition model endorsed by Baker (2001, 2005) and Wexler's (1998) Very Early Parameter Setting model, are compared with Yang's (2002, 2004) Variational model. The Variational model employs statistical learning mechanisms for parameter-setting. Parameter values compete, with delays occurring when the critical input is sparse. Given the uniformity assumption, children in the same linguistic community undergo a similar, gradual development. On the Hierarchical Acquisition model, children initially choose either parameter value, with potential delays arising from hierarchical ordering of parameters. Change is precipitous when initiated. To adjudicate between models, we conducted a longitudinal study of 4 children, ranging from 1;9 to 2;1 at the start of the study, who were in the throes of setting two interlocking parameters governing inflection and negation. Different developmental patterns were observed depending on initial parameter value, and parametric change was precipitous, as anticipated by triggering models.

Keywords: inflection; language acquisition; negation; parameter-setting

1. Introduction

The last thirty years have seen remarkable advances in linguistic theory, and corresponding advances in our understanding of how children acquire language. Advances on both fronts have resulted in large part, in our view, because of a shift from the 1980s rule-based theories of grammar to the current Principles–and–Parameters approach (e.g., Chomsky 1981, 1995). The Principles–and–Parameters approach enabled researchers in language development to make many new and far-reaching predictions about the course of language acquisition. According to this framework, children were no longer expected to accrue indivi-

The research reported in this paper was supported in part by NSF Grant #BCS 9905199 to Rosalind Thornton. The grant also provided doctoral support for Graciela Tesan.

We would like to thank Stephen Crain for extensive discussions of this material. We also thank the audiences at the *ALS Conference* in Brisbane (*Australian Linguistic Society*, July 2006), and the *LSA Annual Meeting* in Anaheim (*Linguistic Society of America*, January 2007) for their comments, and the anonymous reviewers for their input.

ISSN 1450–3417

Biolinguistics 1: 049–098, 2007
http://www.biolinguistics.eu

dual rules for the local language being spoken around them, as in the earlier versions of linguistic theory. The initial state of the language faculty continued to embody universal principles that establish boundary conditions on children's linguistic hypotheses, and children were not expected to deviate from these principles in the course of language development (see, e.g., Crain 1991, Atkinson 1992, Guasti 2002).

In addition to linguistic universals, certain aspects of language variation took on a new look in the Principles–and–Parameters approach. Many differences across languages were encoded in the language faculty as innately specified parameters, where the parameters established (typically binary) choices among linguistic properties of particular natural languages. The introduction of an innately specified system of parameters in Universal Grammar (UG) was motivated by the desire to ensure that language learning was less burdensome for the learner than it would be otherwise (Chomsky 2002). The new look learner is seen as navigating through an innately specified parameter space that is made available by UG; learning is largely replaced by (or reduced to) parameter-setting (cf. Clahsen 1990). This assisted the theory of UG in meeting its overarching goal of "explanatory adequacy," i.e. to explain children's rapid mastery of the grammar of any natural language (Chomsky 1965, 1986).

In the theoretical literature, parameter-setting was originally conceived as being executed by a "triggering" mechanism that resided in the language acquisition device. Each time the mechanism was engaged, it had immediate and far-reaching consequences throughout a learner's grammar. A metaphor for this mechanism was that of a switch — where the learner simply flicked a switch to one setting or the other in response to some triggering experience that was readily observable in the primary linguistic data. The switch metaphor suggested that, at some circumscribed period during the course of development, the setting of a parameter would be decisively triggered, with one value being adopted rather than the other (Hyams 1986, Gibson & Wexler 1994, Fodor 1998, Roeper 1999).

To continue with the metaphor of setting a switch, if the switch was set one way, then the child's grammar took one form, and if the switch was set the other way, the child's grammar took another form. Parameter-setting was seen to set in motion radical changes in children's grammars, for example from a grammar with null subjects to one with overt subjects, or from a grammar without *wh*-movement to one with *wh*-movement, and so on. It was suggested, moreover, that setting a single parameter might induce the introduction of a cluster of properties into children's emerging grammars. The paradigm case was the Null Subject Parameter (cf. Rizzi 1982) studied by Hyams (1986, 1987, 1989). More recently, Snyder (2001) has investigated developmental predictions associated a cluster of related properties in his research on the acquisition of complex predicates and word-formation.

Although parameters were, admittedly, fixed on the basis of input, it was generally assumed that the ambient input sufficed for "early parameter-setting" (see, e.g., Borer & Wexler 1987, Wexler 1998). Nothing in the theory itself prevented parameters from being set early, so if it turned out that they were not set early, then something outside the theory must be responsible for late parameter-setting. Therefore, it was the "null hypothesis" that parameters were set early. Finally, researchers working within the parameter-setting framework assumed that children were initially free to pick either setting, unless a subset problem would arise if one particular setting were adopted, rather than the other. The possibility of "default" settings was available, in principle, but there was no reason to suppose *a priori* that there were default settings. Another view, advanced by Lebeaux (1988), was that children begin with both parameter values operative, with one of them taking priority in response to input from the local language (cf. Yang 2002; see below).

The observation that children could set parameters to either value immediately raised the expectation that children could initially "mis-set" parameters. That is, the learner could initially adopt a value that was inconsistent with the local language. The mismatch would presumably be easily detected, and soon set straight. Still, it could take a child some amount of time to reset a parameter, and during the period of parameter-resetting, the child would be speaking a fragment of a "foreign" language. Therefore, the investigation of children's early productions promised, potentially, to offer empirical support for the parameter-setting approach. On other approaches, the learner was seen to be attempting to match the input, by accruing rules or constructions on the basis of positive examples.

The earliest empirical support for the Principles–and–Parameters approach was one such case of apparent parameter-mis-setting, reported in Hyams 1986. This was a study of young English-speaking children, who were found to produce sentences that lacked overt subjects in their spontaneous speech. Hyams interpreted children's subject omissions as indicating that children had mis-set the "*Pro*-Drop Parameter." The *Pro*-Drop Parameter distinguishes languages that require overt subjects, such as English, from languages that also tolerate covert subjects as well as overt ones, such as Italian. So, child speakers of English who had mis-set the parameter were seen to be speaking a "foreign" language, at least in part. Over the years, there have been a number of other reports of mis-set parameters, where children were found to be projecting parameter values, rather than being directly guided by the input in language development. Hyams (1986, 1987, 1989), Thornton (1990, in press), Becker (2000), and Armon-Lotem *et al.* (2004) all provide empirical data along this line. Of course, children eventually converge on a grammar that is equivalent to that of adult speakers of the local language, so parameter-resetting must be responsive to the input.

Assuming that the input consists solely of positive data, and lacks negative evidence, it is likely that the values of some parameters must be set in a particular order, to ensure that children can always reset parameters, if need be, using positive data. This is the familiar subset condition. The subset condition is that part of the language acquisition device that prevents learners from succumbing to subset problems. A subset problem would arise if the language generated by one setting of the parameter (call it setting A) is a superset of the language generated by the alternative setting (call it setting B). In this case, if the child chose setting A, and it turns out that setting B was correct for the target language, then positive data would not suffice to inform the child of the error, and the child would not converge on the adult grammar. Since children do, in fact, converge on the same grammar as adults, the solution to this problem is to initially set the parameter to setting B. If B is correct for the local language, then B is maintained. If A is the correct setting, then the input will contain linguistic expressions that are generated only on setting A, and the child can use these expressions to reset the parameter to the new value. We will assume that all parameters whose values fall in a subset/superset relation are initially set to the default, subset value (see, e.g., Berwick & Weinberg 1984, Wexler & Manzini 1987 and other papers in Roeper & Williams 1987, and Crain *et al.* 1994). Setting subset problems aside, the picture of language development that emerged in the early days of the Principles–and–Parameters approach was one in which children could freely choose any parameter value, and would quickly be confronted with relevant input if the value they had adopted was incorrect for the local language.

Although nothing in the theory of UG specifies precisely how parameter-setting might unfold in real time, the "null hypothesis" was that parameter-setting (and even parameter-resetting) would take place early in the course of language development, yielding immediate and far-reaching changes from one kind of grammar to another. However, the empirical data have not unequivocally supported early setting of parameters. There are several ways to explain the lack of fit between theory and data. One way for triggering models to explain the recalcitrant data is to invoke performance factors to account for children's unexpected behavior. Another response is to invoke maturation for late-developing grammatical properties (e.g., Borer & Wexler 1987, 1992, Wexler 1994, 1998). Another approach is to say that children scan the data for "cues" or pieces of abstract structure to set parameters, and may delay in setting them if the relevant cues fall below a certain threshold (Lightfoot 1999, also Fodor 1998). A different kind of response to the recalcitrant data is to bring statistical learning mechanisms into play, alongside the principles and parameters of UG. We will scrutinize this last approach, focusing on one important model of parameter-setting augmented by statistical learning, advanced in Yang (2002, 2004).

Yang (2002) contends that the conception of parameter-setting as "triggering" is simply wrong. On Yang's Variational model of parameter-setting, parameters are set on the basis of statistical information contained in the ambient input. On this model, different parameter values amount to different grammars, which compete with each other. The value that survives is the one that is better instantiated in the positive input. There is abundant input for some parameters of course, and the learner is expected to decide on the correct value of such parameters more quickly than when the input is less abundant. A gradual learning curve should be witnessed in both cases, though naturally when the input is abundant, the curve is less gradual. Yang points to evidence of late parameter-setting in support of the Variational model.

In our view, it is premature to cast out the triggering model of parameter-setting in favor of a model that postulates a statistical learning mechanism in addition to UG, even in cases of parameters for which the input is impoverished. The empirical data that have been invoked in support of gradual learning have generally been from children's naturalistic productions, frequently averaged over groups of children and across extended time periods, often months and even years. As a result, these data may not be fine-grained enough to reveal abrupt changes that occur in the grammars of individual children.

To provide richer data sets for individual children, the present study reports longitudinal data that were obtained for four children using elicited production techniques in addition to recordings of naturalistic data. The elicited production studies produced relatively dense data sets for each child subject. These data sets enabled us to accurately track rapid changes in the grammars of the four children whose linguistic progress is studied in this paper. Analysis of the data allows us to draw a picture of grammar formation with sharp contours rather than gradual climbs, as anticipated by triggering models of parameter-setting, and not as expected on the Variational model.

The paper is structured as follows. In section 2 we introduce three models of parameter-setting, and establish a set of criteria by which these models can be distinguished. Sections 3 and 4 discuss the learning trajectory anticipated by triggering models and by the Variational model. A second distinguishing feature of the models, called conformity, is the focus in section 5. In section 6, the models are related to previous literature on children's acquisition of morphosyntactic properties. Two functional parameters from children's developing morpho-syntax are introduced in section 7, and the learnability of these parameters is discussed in section 8. Section 9 presents the details of the study, and section 10 presents the findings of our empirical investigations of the two parameters, and evaluates how well the models stand up against the child language data. Finally, conclusions are presented in section 11.

2. Criteria for Evaluating Models of Parameter-Setting

We will evaluate three theoretical models of parameter-setting, comparing the predictions of these models against findings from detailed investigations of the acquisition of inflection and negation. UG assumes a dominant role in all three of the models. However, the models differ in several important respects. They differ in predictions about:

- (A) the time course of parameter-setting;
- (B) the need for statistical learning mechanisms in parameter-setting;
- (C) how parameter values are engaged, i.e. whether children start with a single parameter value or with both values operative;
- (D) the behavioral patterns that should be observed in parameter-setting, i.e. whether behavior should take the shape of a gradual curve or a steep climb; and
- (E) whether or not the behavior patterns in parameter-setting should assume the same form for all children.

Our joint goals are, first, to spell out the ways in which the three models differ and, then, to see how well each model stands up to empirical findings from longitudinal production studies focusing on the acquisition of morphosyntax in four English-speaking children. The three parameter-setting models are:

(1) *Three Parameter-Setting Models*

i. the Very Early Parameter Setting model (Wexler 1994, 1998)
ii. the Hierarchical Acquisition model (Baker 2001, 2005)
iii. the Variational model (Yang 2002, 2004)

The first two models are similar in character. Both of these models assume that parameter-setting is accomplished without statistical learning mechanisms. However, the Hierarchical Acquisition model introduces an ingredient beyond that of the Very Early Parameter Setting (VEPS) model, namely parameter ordering. Parameter ordering leads to empirical predictions that distinguish the Hierarchical Acquisition model from of the VEPS model. The third model, the Variational model, introduces statistical learning into parameter-setting. The assumption that statistical mechanisms play a critical role in development has taken a strong hold in the field, so it is instructive to explore the proposal that statistical mechanisms are engaged by learners in parameter-setting. To frame discussion of the alternative parameter-setting models, we list a number of criteria by which the predictions of the models will be evaluated using data from child language.

2.1. *Continuity*

The continuity hypothesis maintains that each value of a parameter is fully specified by UG, and that each value corresponds to a fragment of a possible human language (cf. Pinker 1984, Crain 1991, Baker 2001, Crain & Pietroski 2002). According to this hypothesis, at any stage of acquisition children are drawing on properties from a possible human language, but perhaps not using all and only structures exhibited in the local language. The Hierarchical Acquisition and the Variational models assume continuity. By contrast, the VEPS model allows that certain linguistic principles are biologically timed to become operative later than others in the course of development. Before these linguistic operations mature, child grammars may lack certain linguistic properties that characterize adult grammars although they may be latent in UG (cf. Borer & Wexler 1987).

2.2. *Uniformity*

Uniformity is the supposition that all children in the same linguistic community encounter a similar distribution of relevant exemplars (linguistic expressions or structures) for setting parameters. This means that, in the long run, the relative frequencies of the input corresponding to each parameter value are roughly the same for every child. All three models under consideration assume uniformity.

2.3. *Ordering*

Parameter-setting models either postulate that parameters are set in a particular order or that parameters can be set in any order. On the Hierarchical Acquisition model, parameters are hierarchically organized and learners confront parameters in the order imposed by the hierarchy (see also early work on parameter ordering by Nishigauchi & Roeper 1987, Roeper & de Villiers 1991, and more recent approaches in Fodor 1998 and Lightfoot 1999). An ordering of the parameter space could also be imposed by maturation, with certain parameters being biologically timed to become operative at a later point in development than others. Unordered parameters are said to be "independent." Drawing on an analogy in Lasnik & Crain 1985, if parameters are independent, then acquisition is like a scavenger hunt, where items (values) may be acquired in any order. This can be contrasted with a treasure hunt, in which items must be acquired in a particular sequence. The Hierarchical Acquisition model views parameter-setting as a treasure hunt; the Variational model and the VEPS models view it as a scavenger hunt. Without additional assumptions, the scavenger hunt models predict more rapid acquisition (i.e. the completion of parameter-setting) than does a treasure hunt model.

2.4. *Starting Point*

This refers to the number of values that are in play when the learner first engages in setting a parameter. According to the Variational model, the learner entertains multiple values simultaneously (cf. Lebeaux 1988, Valian 1991). On the VEPS model and the Hierarchical Acquisition model, the learner initially adopts a single value of a parameter.

If a single value is selected, there may be a default value or learners may opt for either parameter value, unless this gives rise to subset problems. Default or unmarked values are essential for parameters whose values stand in a subset/ superset relation, on both the VEPS model and in the Hierarchical Acquisition model. Both models assume that, in all other cases, learners are free to select either value as their initial guess.

2.5. *Requisite Input*

One possibility is that the primary linguistic data that suffices to set any parameter is available in sufficient quantity to ensure its "easy" acquisition. This is the position taken by VEPS and the Hierarchical Acquisition model. The Variational model assumes that the learner needs to accumulate a certain amount of data as a prerequisite to setting any parameter, and it contends that the requisite data is not uniformly available for all parameters. On this model, it is more difficult to establish the "correct" value of parameters with sparse relevant input, as compared to parameters that have abundant relevant input.

2.6. *Trajectory*

This refers to the pattern of development that learners manifest in selecting the value of a parameter in response to relevant input. If parameters are set using minimal input, or if input is abundant for all parameters, then no special record keeping is required for parameter-setting. This is the view of VEPS and the Hierarchical Acquisition model. In cases of parameter-resetting, the ("idealized") developmental pattern that is expected is a step function, or rapid incline in one value of the parameter, and a corresponding, and equally rapid decline in the alternative value. Alternatively, record keeping in the form of statistical learning may be required for parameter-setting. This is the perspective of the Variational model.

2.7. *Conformity*

According to this feature of development, either all learners navigate the same course through the parameter space, or children may chart different courses. On the Hierarchical Acquisition model, parameters are ordered, so individual differences may arise, even with uniform and abundant input, as long as children are permitted to adopt different initial values (starting points). Some children will immediately advance through the hierarchical parameter space, others will make just a few missteps, and some children will make many missteps, and will take more time than other children do to complete the process of parameter-setting. On the VEPS model, parameters are set so early that no individual differences will be discernible. The Variational model does not expect individual differences either. If children all start with both parameter values operative at roughly the same rate, parameters are not encountered in a hierarchical ordering, and the input is uniform for all parameters, then individual differences are not expected.

2.8. *Summary*

With these evaluation criteria at the ready, let us briefly summarize the main characteristics of the three models. First, Wexler's (1994, 1998) VEPS model postulates:

- (A^1) parameters are independent (*ordering*);
- (B^1) children initially begin with a single parameter value, but may adopt either value, unless this would lead to subset problems (*starting point, initial value*);
- (C^1) grammar formation is characterized by abrupt changes in grammars (*trajectory*);
- (D^1) differences in the primary linguistic data have little impact on the observed course of parameter-setting (*requisite input*), so no special (e.g., statistical) learning mechanisms are needed to assist in parameter-setting;
- (E^1) since parameter-setting is completed early, little individual variation will be observed (*conformity*).

The VEPS model has little room to maneuver in response to apparent delays in parameter-setting. Maturation is one possibility. Late emergence could also be interpreted as evidence that some phenomenon does not properly count as a parameter. This is the approach taken by the VEPS model for the so-called optional infinitive stage of language development. We return to this in section 6.

The second model is the Variational model (Yang 2002, 2004, Legate & Yang 2007). On this model:

(A^2) parameters are independent of each other (*ordering*);

(B^2) children initially begin with competition among parameter values (*starting point*);

(C^2) grammar formation is characterized by gradual changes in grammar (*trajectory*);

(D^2) differences in the primary linguistic data determine the observed course of parameter-setting (*requisite input*), because stochastic learning mechanisms determine the course of parameter-setting;

(E^2) since input is assumed to be uniform across children, individual differences are not anticipated (*conformity*).

In contrast to VEPS, the Variational model sees the optional infinitive stage of development as falling within its purview. In fact, optionality in children's behavior is probably the principle motivation for the assumption that parameter values initially compete against each other (*starting point*).

The third model is the Hierarchical Acquisition model, based largely on the "implicational universals" proposed in Baker (2001, 2005). On the Hierarchical Acquisition model:

(A^3) parameters are interlocked (*ordering*);

(B^3) children initially begin with a single parameter value, though either value may be selected (*starting point, initial value*);

(C^3) grammar formation is characterized by abrupt changes in grammars (*trajectory*);

(D^3) differences in the primary linguistic data have little impact on the observed course of parameter-setting (*requisite input*), so no special (e.g., statistical) learning mechanisms are invoked in parameter-setting;

(E^3) setting some parameters can only occur once others have been set, and since children may adopt different starting values, different children may set the same parameters at different times (*conformity*), giving rise to individual variation.

On this model UG orders parameters in a hierarchy, with large-scale parameters at the top of the hierarchy, including the Polysynthesis Parameter and the Head Directionality Parameter (cf. Baker 1996). These parameters are presumably set early and have significant impact on the overall form of the language that is acquired. Smaller-scale parameters reside lower in the hierarchy, and they are not necessarily set early because they must await the decisions about parameters that are more dominant in the hierarchy.

The criteria we have elaborated for evaluating the alternative models of parameter-setting should make it straightforward to adjudicate between them, once we turn to the empirical data from child language. For example, all three models anticipate that (at least) some parameters will be set early, but the models differ in expectations about precisely which parameters will be set early. VEPS maintains that all of them will be. Parameters are set early, on the Hierarchical Acquisition model, if they make the broadest cuts across natural languages, so early parameters include the Polysynthesis Parameter, the Head Directionality Parameter, the *Wh*-Movement Parameter, and so on. Other parameters relevant for particular language families will be set later. The Variational model contends that parameters that are associated with the most robust input will be set early. Other criteria will prove valuable in comparing the models, including *trajectory*, to which we turn next.

3. Trajectory: Triggering Models

The trajectory of acquisition data was first used to distinguish competing accounts of grammatical development in the early literature on parameter-setting. The earliest use of trajectory concerned the "*Pro*-Drop Parameter." The *Pro*-Drop Parameter is probably the most thoroughly investigated of all parameters.[1] Given that it governs the use of subjects, and all sentences have subjects, there should be no shortage of data available to establish the time course in setting the parameter, based on children's spontaneous speech. Early research concluded that children learning English initially adopted the [+*pro*-drop] value of the parameter, even in languages in which the adult setting was the [–*pro*-drop] value. This conclusion was based on children's notable omissions of subjects in their spontaneous speech (Guilfoyle 1984, Lillo-Martin 1986, Hyams 1986, 1987, Lebeaux 1987, Jaeggli & Hyams 1988, Pierce 1992, and others). This particular piece of evidence for parameter-setting was challenged, however. One challenge attempted to explain children's omissions of subjects as performance errors, rather than as revealing children's emerging linguistic competence. This position was taken by Paul Bloom (1990) among others (e.g., L. Bloom 1970, Pinker 1984, Valian 1991).

Bloom (1990) proposed that the proportions of null subjects in children's productions could be accounted for by a model of language processing. Using the transcripts of Adam, Eve and Sarah in the CHILDES database (MacWhinney 2000), Bloom (1990) showed that children produced higher proportions of null subjects in sentences with longer VPs than in sentences with medium-length or short VPs. His observation was that in sentences with short VPs children tended to produce more lexical NP subjects, and pronominal subjects tended to appear more often than null subjects in sentences with medium-length VPs.

In response to Bloom's performance account, Hyams & Wexler (1993) provided a number of arguments in favor of an account based on children's linguistic competence. Our discussion is limited to one of their arguments, which rests on the assumption that a competence-based account would be supported by

[1] The exact formulation of the parameter has been much debated, and is not important for our purposes. See Rizzi 2005 for a new approach.

abrupt changes in child language: its trajectory. On the particular version of the parameter-setting theory advanced by Hyams & Wexler, English-speaking children who were omitting subjects were speaking a topic-drop language; thus they had mis-set a parameter. On Hyams & Wexler's analysis, children were expected to use few overt pronominal subjects, because the null pronoun option would be available to them. To such children, null subjects should be the favored option. However, once the parameter was reset to the adult English value, null pronouns should no longer be licensed in children's grammars. Therefore, these researchers predicted a sharp increase in the proportion of overt pronominal subjects once the parameter was reset. But, since null subjects would be replaced by pronominal subjects, no significant change in the proportion of lexical subjects was expected as the parameter was reset; the proportion of lexical subjects should remain constant. The performance model advanced by Bloom (1990) made no predictions about changes in the proportions of null subjects versus pronominal subjects in children's developing grammars; it simply predicted that lexical subjects would tend to be "replaced" by pronouns, or omitted, as processing demands increased, such as in sentences with longer verb phrases.

In assessing the fit of the data to the grammatical model, Hyams & Wexler (1993) turned to the Brown corpus in CHILDES, and investigated eight 2-hour transcripts from the corpora of Adam and of Eve. To measure the overall shift in the proportions of covert subjects versus overt pronouns, they calculated the proportion of lexical subjects and pronominal subjects produced by Adam and Eve in the first and last of the eight transcripts, as well as in a later 9th transcript (cf. Hyams & Wexler 1993: 443). These data are summarized in Table 1. It can be seen that while the use of lexical subjects remained stable over time for the two children, the use of overt pronouns increased by 56% for Adam and by 53% for Eve.

	Adam (2;5-3;5) transcripts 06-30		Eve (1;6-2;3) transcripts 02-20	
	Pronouns	*Lexical subjects*	*Pronouns*	*Lexical subjects*
First transcript	11 %	33%	29%	11%
Last transcript	67%	30%	82%	11%
	56% increase	3% drop	53% increase	no change

Table 1: Proportions of pronouns and lexical subjects in the transcripts of Adam and Eve across 8 transcripts

Hyams & Wexler further remark:

> From the first to the last transcript the proportions of lexical subjects are about the same, and this is true for both Adam (.33 to .30) and Eve (.11 to .11). The proportions of pronouns, however, show a dramatic shift, for both Adam (.11 to .67) and Eve (.29 to .82). Thus, the overall pattern of change from the null subject to the non-null subject stage is a dramatic increase in

> the number of pronominal subjects with a (roughly) steady number of lexical subjects. This is exactly what we would expect under the grammatical model, since null subjects trade off with pronouns under this theory. (Hyams & Wexler 1993: 444)

The "dramatic increase" they noted in Adam and Eve's grammars is impressive, but the figures summarize changes that took place in over a year for Adam, and over 9 months for Eve. These periods are likely to be long enough to be accounted for by models of gradual change, such as the Variational model.

However, closer examination of Hyams & Wexler's data reveals that the dramatic increase in pronominal subjects actually took place within a much shorter time period. To show this, we present graphs of the transcript-by-transcript data for each child from Hyams & Wexler's Table 4 (Hyams & Wexler 1993: 443). The data in Figures 1 and 2 below show the proportion of lexical subjects and overt pronominal subjects produced by each child in each session. The proportion of null subjects is calculated by adding the overt pronouns and lexical subjects together and then subtracting the sum from 100. Because lexical subjects remain stable over time, as null subjects decrease, there is a corresponding increase in pronominal subjects.

Figure 1 shows the graph of Adam's data. A dramatic change takes place between transcripts 14 and 20 (ages 2;9.18 and 3;0.11). At transcript 14, null subjects are produced 70% of the time; by transcript 20, they have dropped to 12%, a change of 58%. At transcript 14, overt pronominal subjects appear only 15% of the time; at transcript 20, they comprise 77% of Adam's subjects, an increase of 62%. Thus the dramatic change in use of pronominal subjects takes place within 3 months.

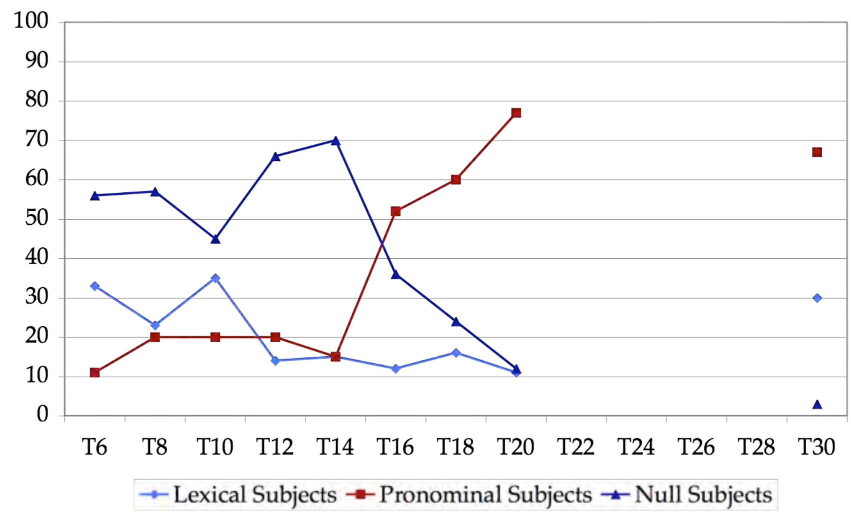


Figure 1: Adam's subjects from transcript 06 to transcript 20, and 30. Ages 2;5.12 to 3;0.11 and age 3;5.1

The data for Eve are illustrated in Figure 2. Eve's null subjects are replaced by pronominal subjects during the period from transcript 2 (at age 1;6.1), where null subjects comprise 60% to transcript 12 where null subjects comprise only 11% of the total. At that point, Eve is 1;11. Thus, in 5 months null subjects have decreased by 49%, and pronominal subjects have increased by 39%.

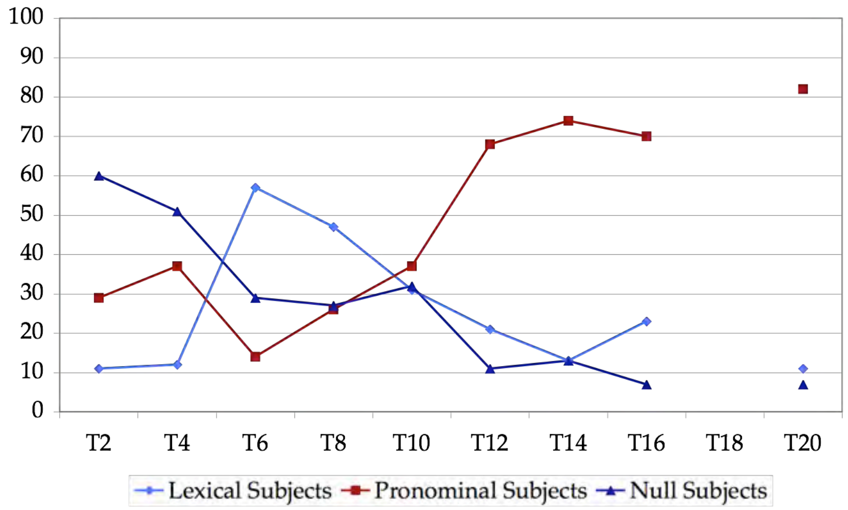


Figure 2: Eve's subjects from transcript 02 to transcript 16, and then 20. Ages 1;6. to 2;1 and 2;3

The shifts that take place from transcript 14 to transcript 20 for Adam, and from transcript 2 to transcript 12 for Eve, resemble the pattern of responses that appear in studies of "categorical perception." Apparently, one type of structure (roughly, one category) is completely replaced by another as some perceptual feature (here, time), is manipulated. An idealized depiction of what we will call "categorical acquisition" appears in Figure 3. This is the trajectory pattern that is expected on "triggering" models such as the VEPS model and the Hierarchical Acquisition model.

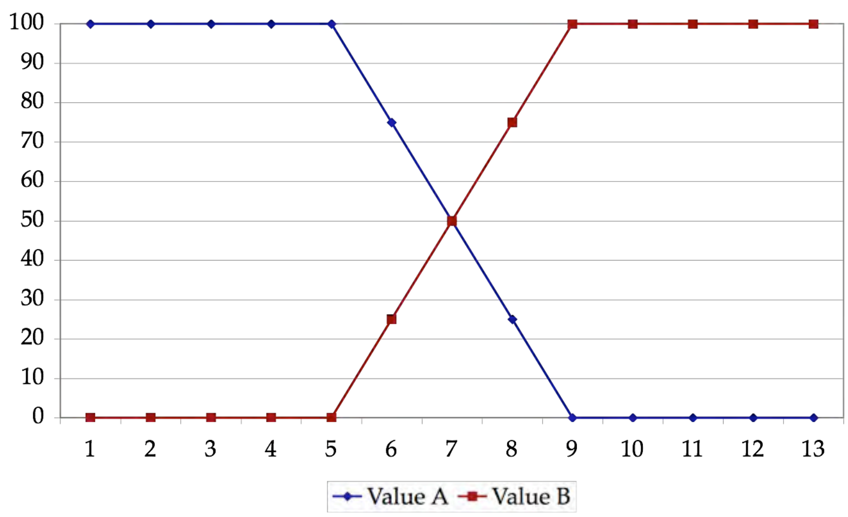


Figure 3: Idealized trajectory of categorical acquisition

In principle, categorical acquisition occurs when one sentence type, based on the initial setting of the parameter, is used in 100% of children's productions, and then drops to 0% once the parameter is reset. This makes no allowance for noise, however. To accommodate noise in child language, we simply adopt the standard criterion (e.g., Brown 1973) that 90% "correct" adult-like usage in obligatory contexts indicates that a sentence structure has been acquired. This means that an abrupt drop from (at least) 90% consistent usage of one kind of structure, to 10% (or less) consistent usage is evidence of a categorical transition from one value of a parameter to another. At the same time, the structure associated with the "new" parameter value should have increased from 10% consistent usage to 90% during the same period of time. And, for the transition to be categorical, grammatical change must occur within a confined timeframe. There is no standard criterion for this aspect of categorical acquisition established in the acquisition literature. As a first pass, we suggest that the transition from one value to another should occur within a three-month period, following which the non-adult value should not be exemplified more than 10% of the time.[2]

In practice, these criteria may run into practical complications, for example where one structure does not simply replace an alternative structure. The case of null subjects discussed by Hyams & Wexler (1993) is one such example. While subject omissions disappear almost completely from children's productions (constituting only 3-7% of children's productions), the emergence of overt pronouns does not reach 90% consistent usage, because lexical subjects are another option. In short, when optional sentence structures complicate the picture, changes in proportion of consistent usage may not be as dramatic as in Figure 3. Exactly what increase should be counted as categorical acquisition depends on the phenomenon being investigated. In the example of Adam and Eve's development, evidence of parameter-setting on a "triggering" model consists of an over 50% increase in usage of the structure associated with a new parameter value, i.e. pronominal subjects.

4. Trajectory: The Variational Model

What course of acquisition is expected on a statistical learning model of parameter-setting, such as the Variational model? This model supposes that children initially attempt to parse the linguistic input using two "grammars," one with each value of the parameter operative in it. If one of these competing grammars parses the input successfully, that grammar is "reinforced," increasing the probability that it will be used in the future. Assuming that the grammar with the alternative parameter value is unable to parse the same input, then that grammar is "penalized," and its probability of being selected in the future is correspondingly reduced. Gradually, probability weights are adjusted until the grammar with the non-target parameter value is no longer a contender, and becomes obsolete.

[2] Recall that Eve's change takes place in 5 months, rather than 3 months, but it should be noted that she is considerably younger than Adam. It may be that in the future, another consideration will be the age of the child at the time of setting the parameter.

On this model, quantitative data from input frequencies can be used to estimate the learning trajectory, that is, whether a parameter setting will be consolidated early or late. In support of this proposal, Yang (2002) draws on the findings from the literature on child language. Reports from the literature suggest, for example, that French speaking children learn that French is a verb raising language by 1;8, (based on data from Pierce 1989). The sentences that provide the informative data about which parameter setting is correct are called "signature" sentences by Yang. For verb raising, the signature sentences are of the form VFIN Neg/Adv. Using transcriptions of adult speech to children in the CHILDES database, Yang estimates that the VFIN Neg/Adv signature sentences make up 7% of all sentences children acquiring French hear. Thus, Yang concludes, the frequency of signature input of an early set parameter must constitute at least 7% of the input data. On the other hand, if the requisite sentences are rare in the input data for some parameter, the Variational model would be forced to predict that the parameter would be set relatively late in the course of development. Drawing on Valian's (1991) summary of child data, which reveals null subjects not disappearing from children's productions until about 3 years of age, Yang (2002) concludes that the signature sentences must be rare. Yang assumes that the requisite input consists of sentences with expletive *there* subjects; such sentences cannot be parsed by the setting of the parameter that licenses null subjects. Yang's counts from the CHILDES database estimate that expletive *there* sentences comprise 1.2% of the adult input to children. Thus, as a working hypothesis, Yang proposes that there will be late parameter-setting if the signature sentences comprise 1.2% or less of the input to children. The quantitative predictions are matched with further empirical data in Yang (2004). These data include Thornton's (1990) observation that some children ask non-adult long-distance *wh*-questions with a copy of the question word, such as *What do you think what pigs eat?* until 4 or 5 years of age. The findings are accurately modeled on the statistical learning account, because adult long-distance *wh*-questions constitute only 0.2% of the input data. In short, the speed with which parameter values are adjusted in child grammars depends on the character of the input, according to the Variational model.

Depending on the relative frequency of signature sentences, one parameter value may rise more quickly to dominance, or there may be a prolonged struggle between the two values. The main point is that gradualness is expected in the rise and fall of many competing parameter values, rather than abrupt changes. This scenario contrasts with rapid ascent and descent of parameter values that is always expected on a triggering model, when parameters are switched from one value to the other. An idealized trajectory, based on the statistical learning implemented in the Variational model, is presented in Figure 4.

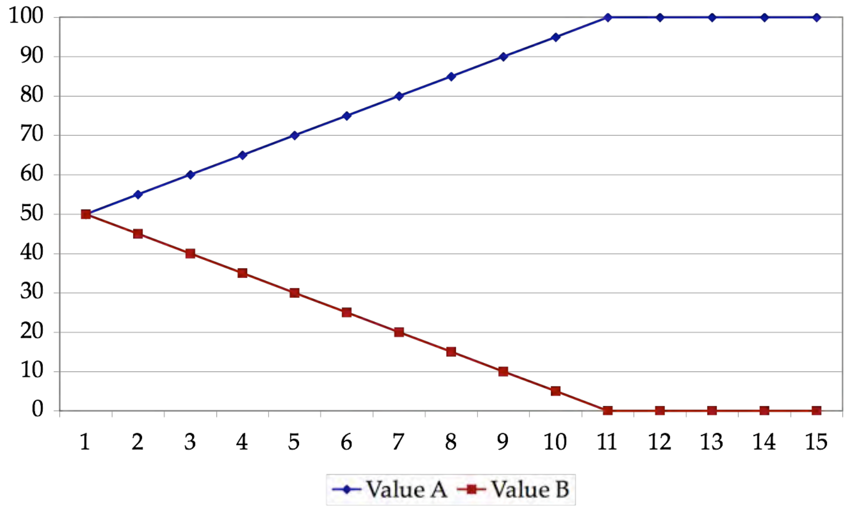


Figure 4: The trajectory of parameter-setting on the Variational model

5. Conformity

Conformity is another criterion that will be invoked to distinguish among the alternative parameter-setting models. The Variational model expects a similar developmental pattern for all children. The expectation of conformity is derived from two sources. First, it is assumed that every child is exposed to the same distribution of structures in the ambient input, so the signature data for the target parameter value should appear in similar proportions for every child. Second, although children access the two parameter values probabilistically, the expectation is that most children will access the two values at around 50–50. Put another way, it is highly unlikely that a child would access one value 100% of the time, and the other value not at all (as in the triggering model). Taken together, these features of model, and the input, should conspire to make every child display a similar learning curve.

The model can be augmented with a "learning parameter" that adjusts how much penalty or reward should be awarded to grammars for success or failure in parsing the input data, thus potentially speeding up the rate of learning as the learner accumulates more data and becomes more confident. This would mean that the learning curve would look more like the curve seen when parameter-resetting is triggered, with more rapid change nearer the time of convergence on the adult grammar. However, Yang notes that to implement such a changing learning rate is "computationally expensive" and that it alters the mathematical properties of the proposed model. He also adds that such approaches "deviate from the guidelines of psychological plausibility and explanatory continuity that acquisition models are advised to follow" (Yang 2002: 49). However, the idea that children with Specific Language Impairment, as a group, have a different value for the learning parameter than children acquiring language normally is entertained in Legate & Yang (2007), but there is no suggestion that the learning rate for any given parameter varies across individual children. In sum, the expectation of the Variational model is that every child should display a similar learning curve.[3]

Conformity is not expected on the Hierarchical Acquisition model envisioned in Baker 2001, 2005. On this model, parameters are ordered, or at least partially ordered. At the top of the hierarchy is the parameter that draws a typological division between polysynthetic languages and the others. This parameter determines a range of parameters on each side of the hierarchy for the learner's subsequent consideration. As the learner traverses one side of the hierarchy or the other, the parameters that are subsequently encountered differentiate among fewer and fewer languages. On the Hierarchical Acquisition model, delays in parameter-setting are a logical consequence of the structure of the hypothesis space. As Baker remarks, "an efficient learner should learn in a structured way in which some parameters are entertained first and others later" (Baker 2005: 95). The broad typological parameters at the top of the hierarchy could even be set before the onset of production, as Wexler (1998) claims. Obviously, this is not necessarily true of parameters that are situated towards the bottom of the hierarchy. These minor parameters are seen to be fine-grained features of the local language, with the relevant parameters being set, presumably, after children begin to talk, and possibly much later than that (cf. Baker 2001: 192-195).

As noted earlier, the Hierarchical Acquisition model also accommodates parameter mis-setting. As a triggering model, children initially begin with a single parameter value, though either value can be selected. The model does not prevent the child learner who initially selects the wrong value for a parameter from stalling briefly at a particular point in the hierarchy as further input data is assessed, such that the parameter can be reset. However, there is no assumption of statistical learning, so the model anticipates that the trajectory should take the form of categorical acquisition. Although the hierarchy minimizes the burden of learning, wrong turns are not eradicated completely, and so the model allows for discrepancies between children in the timing of parameter-setting. Thus the Hierarchical Acquisition model does not require conformity of children.

The next section turns to the literature on children's acquisition of morphosyntactic properties, and examines why Wexler considers children's development to be outside the purview of parameter-setting, whereas Yang embraces parameter-setting as an explanation of optionality in child language.

6. Early Morphosyntax

Young children's developing knowledge of inflection across a diverse number of

[3] There may be cases where it is reasonable to accommodate individual variation with statistical learning, but this means abandoning uniformity. Input related to certain linguistic structures that lie at the periphery, such as use of metaphors and idioms, or semi-productive structures (like the *time away*-construction mentioned in Goldberg 2003) might differ across individuals. On the other hand, it is unlikely that such peripheral constructions are used for setting parameters. The parameters studied in this paper are concerned with "core" grammar — the functional categories of inflection and negation. These morphosyntactic properties are unlikely to differ significantly across speakers.

languages has been the subject of intense scrutiny in the last 15 years. The impetus for this research program was the observation from Romance and Germanic languages that, in the earliest stages of acquiring language, children sometimes fail to use inflected verbal forms in matrix sentences, as adults do, but permit the infinitival form of the verb. This occurs in situations when the intended interpretation clearly refers to an event in the present or in the past (e.g., Pierce 1992, Poeppel & Wexler 1993, Wijnen 1997, Hyams & Hoekstra 1998, Hyams 2007). This phenomenon has been called the "optional infinitive" stage of language development by Wexler (1994, 1998) and the "root infinitive" stage by Rizzi (1993, 2005), respectively.

In many language families, the utterances from child language that are non-finite are readily identifiable because the verb has the morphological form reserved for the infinitive. English stands apart from Romance, Germanic, and Slavic languages in this regard, because infinitives bear no special morphology. Rather, the verb form used for the infinitive in English is just the verb stem preceded by *to*. Such "*to*+stem"–forms do not appear in early child English, however (Goro 2004). Instead, 2-year-old English-speaking children use the verb stem alone. Instances abound of children producing utterances like *Daddy go* instead of *Daddy goes* or even *Daddy to go.* Children's failure to use appropriate morphology on verbs that should express tense and/or agreement was interpreted by Wexler (1994, 1998) to be the English instantiation of the optional infinitive phenomenon, and this interpretation of English-speaking children's data has been generally accepted. The range of research establishing the properties of English optional infinitive utterances is extensive and we cannot hope to review it all here. See Guasti 2002 for a comprehensive summary of the literature.

During the optional infinitive stage, English-speaking children "optionally" produce utterances with no tense or agreement. In many children, this stage lasts until the child is 2 and a half or even 3 years old. So, the behavioral profile of this stage does not accord with very early parameter-setting, and is not considered to be within the purview of VEPS. In various proposals, over the years, Wexler and his colleagues have argued that tense and agreement (or the mechanisms that govern them) are in some way deficient in young children's grammar.

One example is the Agreement Tense Omission Model proposed by Schütze & Wexler (1996). This model contends that young children often fail to project either tense and/or agreement features in a syntactic derivation. Together with the assumption that nominative case is assigned by agreement, Schütze & Wexler make a number of predictions about the combinations of subject and verb that are licensed in child grammars at the optional infinitive stage. If the child assigns both tense and agreement features, as in the adult grammar, the 3rd person agreement marker will correctly appear on the verb. But if the child fails to assign the agreement feature in the course of the derivation, default case (i.e. accusative case) will appear on the subject (if it's a pronoun) and the verb will lack appropriate morphology, as in *Him cry*. Alternatively, if agreement is in place, but the tense feature is lacking, pronouns will manifest nominative case, but the 3rd person morphology will be omitted, as in *He cry*. Still other properties of children's syntax are seen to follow from the optionality of tense and agreement features in the representation at this stage of language growth. As markers of tense and agreement, auxiliary verbs are often missing, and children do not use *do*-support.[4]

For our purposes, it is useful to contrast the Agreement Tense Omission Model with the Variational model. The Variational model accommodates the optional infinitive stage of language development because, on this model, children initially hypothesize dual grammars (or, equivalently, dual values for each parameter). Each value of a parameter begins with a roughly 50–50 chance of "success" at the start. As input is encountered, the weights of the alternative values are adjusted; the value of the parameter that fails to parse the input is punished and devalued, thereby indirectly favoring the alternative value. During the optional infinitive stage of language development, English-speaking children are seen to be vacillating in the use of tensed and non-finite forms of a verb, according to the Variational model. Children eventually encounter more evidence that verbs require tense in English. But this takes time because, as Yang calculates, there are only 8% more unambiguous finite verb forms than forms that are ambiguous in marking finiteness. Although the fluctuating early utterances may cause children's productions to look random, in fact the child is simply instantiating the various parameter values that underlie natural languages. Children's non-adult utterances are, therefore, completely consistent with UG, and in keeping with the continuity hypothesis. In this respect, the Variational model is in agreement with the Hierarchical Acquisition model, where children are also seen to hypothesize parameter values that represent properties of other adult languages (e.g., Crain & Pietroski 2002).

On the Variational model, UG provides children with a parameter that categorizes languages into ones that exhibit overt tense morphology versus ones that do not. Languages in which infinitives are observed in matrix clauses are also languages that express tense overtly. Still other languages lack tense morphology, such as Chinese.[5] The child's task is to figure out if the local language expresses tense overtly or not. On the Variational model, the optional infinitive stage of child language is a manifestation of the gradual parameter-setting process. The advantage of such a parametric account is that it maintains continuity between child and adult grammars, whereas the VEPS proposal violates the continuity hypothesis.

In support of the Variational model, Legate & Yang (2007) cite data demonstrating variation in the optional infinitive stage in three languages. In Spanish, the relevant phenomenon barely surfaces. Optional infinitives appear in Spanish-speaking children's speech in about 10% of verbs before 2 years of age, and this drops to below 5% by age 2;6 (Grinstead 1994). Children learning French produce optional infinitives more often. Between 15 and 30% were reported for

---

4 More recently, Wexler (1998) has recast the model in more Minimalist terms. We cannot do justice to the model here. The main idea is that the child is unable to carry out feature-checking as adults do, due to a developmental constraint, and this results in the child's production of optional infinitives. Once the child's system of constraints mature, the verbal morphology becomes adult-like. Thus, both Schütze & Wexler (1996) and Wexler (1998) claim children's early syntax must "grow" into adult syntax.

5 Chinese has morphology (particles) expressing aspect, but this is considered to be a separate issue.

three French-speaking children between 1;6 and 2;6 (Rasetti 2003). Finally, following Phillips (1995), Legate & Yang observe that Adam produces a considerable number of optional infinitives at 3;0 and even Eve, whose linguistic precociousness is legendary, was still missing verbal morphology from 50% of her utterances at 2;3, when her recordings stop.

Drawing on the Variational model developed in Yang 2002, Legate & Yang predict a positive correlation between the "informativeness" of the input and the duration of the optional infinitive stage in a particular language. For example, fewer verbs express explicit tense morphology in English as compared to Romance languages. It is therefore anticipated that English-speaking children will take longer to "set" the relevant parameter than children acquiring Romance languages. Searches of input to children from the CHILDES database support this prediction. In Spanish, adult input expresses tense 80.1% of the time, and does not express tense 19.9% of the time. This means that 60.2% of the input is informative about the setting of the parameter, giving the child "ample opportunity to learn that their language makes use of tense" (Legate & Yang 2007: 330). Likewise, adult input in French marks tense on 69.8% of the verbs, and tense is lacking on 30.2% of verbs, yielding 39.6% informative input. English-speaking children have a less informative input. Based on transcripts of adult input to Adam, Eve, and Sarah, Legate & Yang calculate that 52.9% of adult sentences express tense, and 47.1% of adult sentences do not. Therefore the [+Tense] setting of the parameter has only a 5.8% advantage over the competing value. This low proportion of useful data is used to explain the prolonged optional infinitive stage experienced by English-speaking children, as compared to French- and Spanish-speaking children.

To recap, Legate & Yang offer an account of the cross-linguistic variability in the optional infinitive stage that is consistent with the continuity hypothesis. It should be kept in mind that the Variational model entails conformity across children, since children are seen to be assigning weights to the different values for parameters in response to uniform data sets. In the next section, we introduce the two parameters that we will use to evaluate the different parameter-setting models of children's development of morphosyntax. Then we turn to the laboratory, where we describe a longitudinal study of the trajectory and developmental path of these parameters by four children. At that time we will ask about conformity across children.[6]

[6] The optional infinitive stage of child language has proven to be a huge research enterprise, and we cannot review the full range of approaches in this paper. Moreover, in view of the parameters we investigate (one on inflection, one on negation), the remainder of the paper concentrates on children's productions of finite utterances. We would not analyze the optional infinitive stage of child language as an instance of maturation or of parameter-setting, however.

In our view, optional infinitives are produced by children when they are either unsure of, or cannot implement, a parameter setting. We will demonstrate this below, using children's sentences with negation in the period that precedes *do*-support. Following Tesan (2005), optional infinitives can be analyzed as adult-like derivations in which the agreement morpheme is not realized at Spell-Out, due to a deletion repair mechanism (Lasnik 2001). As such, optional infinitives are expected to disappear from children's speech once the parameters for inflection and negation are set. This prediction is upheld in the data we have gathered, but this must remain a topic for another paper.

Three different models of parameter-setting have been introduced (which were first presented in (1) above). From this point forward, our goal is to explain the source of certain non-adult properties of children's speech. Because the VEPS model does not anticipate the kind of data we discuss, our focus will be on the Variational model and the Hierarchical Acquisition model. Like Legate & Yang (2007), we propose that children's morphosyntactic behavior is governed by parametric decisions. We turn next to the two parameters that are the focus of our investigation of children's verbal morphology.

7. Two Functional Parameters

The parameters that we propose to investigate are both associated with functional categories, one with inflection (in particular, 3rd person agreement), and the other with negation (cf. Borer 1984, Chomsky 1995). The two parameters both express ways in which languages vary, and for this reason we consider them to be parameters. We readily concede that the properties governed by these parameters may ultimately be recast as the product of "deeper" parameters.[7] We will simply assume, here, that the expression of the parameters is sufficient for the purposes of comparing the parameter-setting models.

The two parameters that we present are relevant for language learners who are traversing the non-polysynthetic side of the parameter hierarchy.[8] Because the two parameters govern functional categories, these parameters sit lower in the hierarchy than do more general parameters such as the Polysynthesis Parameter, the Head Directionality Parameter, the Optional Polysynthesis Parameter, and the Verb Attraction Parameter (cf. Rizzi 2005). Therefore, these parameters are expected to be set later in the course of acquisition that the higher-level parameters, according to the Hierarchical Acquisition model. Administering criteria proposed by Baker to distinguish parameters that are ordered from ones that are not, it turns out that the two parameters we propose are unordered with respect to each other and, therefore, sit side-by-side in the parameter hierarchy.[9]

The Inflection Parameter is based on Lasnik's (1995, 1999) hybrid theory of verbal morphology. On this theory, languages select an inflection category that has the property of being *featural* or *affixal*. This choice between *featural* versus *affixal* determines, for each language, how the verb and its morphology combine in a derivation. Lasnik (1995) does not explicitly call this cross-linguistic difference a parameter, but it lends itself readily to this analysis (cf. Boeckx 2006 on the "Inflection-Attachment Parameter"). On Lasnik's account, if a language selects an inflection category with featural properties (i.e. "uninterpretable" features), then these features are generated in the inflection node in a derivation, and must be checked off by an appropriate category as the derivation proceeds.

---

7 Thanks to a reviewer for making this suggestion.

8 It is possible that polysynthetic languages allow no choice for these morphosyntactic properties; for example, it may be that negation must be a head in these languages.

9 According to Baker, "a parameter Y is subordinate to another parameter X if and only if Y influences just one of the languages types defined by X" (Baker 2005: 123).

Typically, the main verb is the lexical item that raises to check off the uninterpretable features in INFL. So in French, for example, the main verb is inserted into the syntactic derivation already fully inflected, and it moves out of the verb phrase to INFL to check off its uninterpretable features.

Languages that select the affixal value of the Inflection Parameter have different requirements. In this case, affixal features are generated in the INFL node, and the affix (such as the 3rd person *–s*) is inserted into this position. The affixal features generated in INFL have no syntactic requirement, however, so no movement takes place in the syntactic computation. Before the sentence is pronounced, however, the affix must join with a verb so that it is not a "stray," unattached affix (cf. Lasnik 1981). The literature argues that the affix lowers onto the verb at PF (Bobaljik 1994, Lasnik 1999).[10] In the case of present tense/ agreement, the lowering operation is visible only for the 3rd person morpheme; for morphemes associated with other persons, it is assumed that a zero (silent) morpheme lowers onto the verb. English does not manifest uniform behavior, however. Some verbs represent exceptions to the affixal setting of the Inflection Parameter. For example, auxiliary verbs and modals select a featural value for INFL. As in French, auxiliary verbs and modals are inserted into a sentence derivation already adorned with their morphology, and they raise to INFL to check off the uninterpretable features. The language learner thus has to acquire these particular verbal items as selecting a featural INFL.

The second parameter in the acquisition of verbal morphology concerns negation. For simplicity, the parameter is expressed as a binary choice for the syntactic status of negation; the item is either a head or a specifier (cf. Ouhalla 1990).[11] The choice between these options has an effect on potential word orders in the language. For example, in English, when *never* is positioned in specifier position, the *–s* affix can lower to the verb uninhibited. But if the negative item is generated in the head position, it has the potential to block affixation. Thus although the Inflection Parameter and the Negation Parameter are separate parameters, they interact closely, with the Negation Parameter having some effect on the way in which inflection is expressed. In this paper, we limit our investigation of inflection to the 3rd person agreement morpheme.

It is worth reviewing some representative examples of the alternative values for the Inflection Parameter and for the Negation Parameter, too see how they play out in adult languages. The four possible options that the two parameters yield are illustrated in Figure 5. In section 8, we explore how these options might play out in the grammars of children learning English.

10 The parameter is closely related to the Verb Raising Parameter, but it is considered to be independent of it.

11 For a different formulation of the parameter, see Tesan 2005, where the binary settings of the Negation Parameter are considered to be *affixal* and *featural*.

Inflection–Negation

Featural Neg–Spec	Featural Neg–Head	Affixal Neg–Spec	Affixal Neg–Head
French	Spanish	Swedish	English

Figure 5: The two functional parameters

Spanish selects a featural setting for the Inflection Parameter, and negation is generated in head position. In Spanish, the negative lexical item, a head, is not independent, but raises along with the verb to INFL, as shown in (2). (The origin of elements that have been moved is indicated by strikethrough.)

(2) [$_{\text{IP}}$ Juan no habla [$_{\text{NEG}}$ ~~no~~ [$_{\text{VP}}$ ~~habla~~ italiano]]] *Spanish*
Juan NEG *speak.3*SG *Italian*
'Juan doesn't speak Italian.'

French also selects a featural setting of the Inflection Parameter. Like Spanish, the weak clitic form of negation (*ne*) is a head, and raises with the verb to INFL, but this form is often omitted in colloquial language. A second form of negation, namely the lexical item *pas*, is obligatory in negative sentences and is positioned in the specifier position. The example in (3) illustrates a sample derivation. The main verb raises to pick up the negative element *ne* in the head position, and bypasses *pas* in the specifier position as it raises to check its uninterpretable features in INFL.

(3) [$_{\text{IP}}$ Jean ne–parle [$_{\text{NEG}}$ pas ~~ne~~ [$_{\text{VP}}$ ~~parle~~ grec]]] *Spanish*
Jean NEG*–speak* *not* *Greek*
'Jean doesn't speak Greek.'

Swedish also positions negation in the specifier position, as in French. However, unlike French, Swedish selects affixal morphology. Therefore, the verbal affix in Swedish lowers over the negative item *inte* to merge with the main verb. This is most transparent in embedded clauses, where the V2 operation does not mask the operation of affixation. The following example, cited in Tesan (2005) illustrates the word order of Swedish in embedded sentences.

(4) … att Lena inte köpte en ny bok igår. *Swedish*
… that Lena not bought a new book yesterday
'… that Lena didn't buy a new book yesterday.'
(adapted from Vikner 1995: 45)

The fourth combination of negation and inflection is represented by English. English selects affixal morphology, and the negative items *not* and *n't* are heads. Before outlining their behavior, however, it is worth considering the negative adverb *never*. Unlike *not* and *n't*, *never* is positioned in the specifier position of the negation projection; therefore it functions much like *pas* in French and *inte* in Swedish. Since *never* is a specifier, the verbal affix can lower across *never* to the verb, yielding sentences like *He never speaks French*:

(5) [$_{IP}$ he ~~–s~~ [$_{NEG}$ never [$_{VP}$ speak–s French]]]

In contrast to *never*, however, the lexical item *not* is usually analyzed as a head (cf. Chomsky & Lasnik 1991). Consequently, an equivalent utterance like *He not speaks French* is ruled out because *not* blocks lowering of the affix onto the verb. This predicament calls for a quirky rescue operation — "*do*-support." To save the sentence derivation from crashing, *do* is inserted, to provide a host for the stranded affix. This is how the acceptable sentence *He does not speak French* is derived.

(6) [$_{IP}$ he doe –s [$_{NEG}$ not [$_{VP}$ speak French]]]

The most common form of negation in English, however, is *n't*, which is assigned the status of an affix (Zwicky & Pullum 1983). The negative affix *n't* must join with a host auxiliary verb or modal (e.g., *can't*, *haven't*, *isn't*, etc.). Since affixes are heads that attach to other heads, these modals and auxiliary verbs can provide the information that *n't* is an affix.

(7) [$_{IP}$ he doe –s –n't [$_{NEG}$ ~~n't~~ [$_{VP}$ speak French]]]

The next step is to investigate how these four parametric options might be reflected in English-speaking children's grammars.

8. Four Parametric Options in English

On the Hierarchical Acquisition model, learning is guided by the architecture of the parameter hierarchy; the child is led through a structured set of options. If relevant evidence for the target parameter setting is not immediately forthcoming, however, children may hazard a guess, and pick the non-target value. In cases where two unordered parameters sit side-by-side on the parameter hierarchy, as we propose for the Inflection Parameters and the Negation Parameter, children are faced with four options, only one of which matches the local language. Thus, in principle, children have only a 25% chance of picking the right combination of parameter settings. In this section, we flesh out what sort of English utterances would arise on the different options, and what positive evidence children would need to jettison wrong parameter values.

8.1. Learnability

Both the Hierarchical Acquisition model and the Variational model assume uniformity — i.e. that requisite data are available in the input in similar distributions for all children, to ensure that parameters are set to the adult value. The Variational model is concerned with the statistical frequency of the data that brings about parametric change, whereas the Hierarchical Acquisition model has nothing to say on this matter. In the case of English morphology (which requires *do*-support, an unusual operation, cross-linguistically), it appears that the character of the input may have some influence on how parameter-setting takes place. For the Variational model, if decisive input is not readily available, then this entails prolonged acquisition for all children (because of the assumption of uniformity). For the Hierarchical Acquisition model, ambiguous input means that children may have to guess which value of the Inflection Parameter to take as the initial setting. This does not entail late acquisition for all children, since children have a 50–50 chance of selecting the value that is consistent with the local language. So, the Hierarchical Acquisition model anticipates slightly delayed parameter-setting (awaiting decisions about higher level parameters), but prolonged acquisition is anticipated only for children who initially choose the wrong parameter value. So, conformity across children is not anticipated on this model for parameters that are associated with ambiguous input.

Consider now, the input that English-speaking children evaluate in trying to determine the English value for the Inflection Parameter. Affirmative sentences do not provide unambiguous data that confirm that English takes affixal morphology. A sentence like *John speaks French* is ambiguous as to the nature of inflection: Either the verb has raised, in which case the inflection category is featural, or the verb has affixal morphology, and has not raised — there is no way to tell. So, the learner must look elsewhere. Unequivocal data showing that English takes the affixal value of the parameter is presented in sentences with *do*-support, where the form of *do* is 3rd person: *does* or *doesn't*. From such examples, the learner can infer that the 3rd person *–s* morpheme is generated higher than the main verb and requires a morphological host other than the main verb. The observation that *do* is inserted to host the *–s* morpheme informs the learner that English opts for the affixal value of the Inflection Parameter.

For the Negation Parameter, the fact that *n't* appears attached to modals and auxiliaries (*can't*, *shouldn't*, *haven't*, *isn't*, etc.) suggests that it is a head. However, the knowledge that *n't* is a head doesn't help children implement this value of the Negation Parameter in sentences with main verbs. Children must be exposed to the specific lexical item *doesn't* to see that *n't* remains outside the verb phrase, with the *–s* affix positioned higher than negation. So, in principle, *n't* attached to any modal or auxiliary provides the data for setting the Negation Parameter but, in actual fact, it may not be set until children discover *doesn't*. Of course, children's discovery that *n't* is a head still does not guarantee that they recategorize the negative morpheme *not* as a head also; it could remain a specifier. Therefore, children could use *doesn't* in the same way as adults do but at the same time, they could analyze *not* as a specifier. Our empirical findings suggest that once children acquire *doesn't*, they cease to use *not*, at least for a time. For now, we will simply assume that *doesn't* is the critical data that children need to inform them of the target parameter value, and leave the continuing status of *not* in children's grammars as an open question.

Having established that *does* and *doesn't* constitute unambiguous data for learners to set the Inflection Parameter and the Negation Parameter, we can use the frequency of occurrence of these lexical items in the input to estimate whether these parameters are acquired early or late, according to the Variational model. To obtain an estimate of the frequency of these "signature" inputs, we conducted a search of the adult input to Adam and Eve in the CHILDES database (MacWhinney 2000). Of the 30,882 adult utterances that were checked, 466 (1.5%) contained *does* and 296 (0.95%) contained *doesn't*. Both of these items suffice for children to set the Inflection Parameter to the affixal value. If the two figures are combined, then, there are 762 informative instances, which is 2.46% of the total utterances. On the Variational model proposed by Yang, parameters whose signature input appear with a frequency of occurrence of less than 7%, such as the Inflection Parameter, are expected to be set late in the course of acquisition.

Turning to the Negation Parameter, if we assume that any modal, or auxiliary with the *n't* affix is signature input for the parameter, then there are 3,618 relevant utterances in the input, out of total set of 30,882 utterances. This amounts to 12% of the input to these children. So, according to Yang's criterion, the Negation Parameter could be set early. However, as we saw, children need to witness *doesn't* in the input to see how negation is analyzed with main verbs, rather than with auxiliary verbs. If the relevant data is narrowed just to *doesn't*, then there are only 296 relevant utterances: 0.95% of the input. The prediction, again, would be late parameter-setting.[12]

8.2. Child English

On both the Variational model and the Hierarchical Acquisition model, in principle, all four options from the diagram in Figure 3 could be instantiated in children's grammars. On the Variational model, all four options would initially vie for dominance in the learner's grammars, while on the Hierarchical Acquisition model, just one option is expected to be contemplated at a time. These expectations for the two models only hold, of course, if there is a viable way to express these parameter values in English.

Let us begin by considering children's potential utterances, should they hypothesize that INFL has the featural value of the Inflection Parameter. If the children learning English select the featural value, they will succeed in producing adult-like utterances with auxiliary verbs and modals, because these verbal

[12] The frequencies were calculated as follows. Using CLANX, all utterances in the adult tiers of Adam and Eve's files (i.e. MOT, FAT, COL, RIC, and URS) were gathered into a file. A number of utterance types were eliminated from the data set, including fragments of various kinds (NP, AP, PP, and other non-sentential utterances), and any utterance that contained 'xxx' in the transcription. In the 30,882 utterances that remained, we searched for any occurrence of *does/doesn't*, including occurrences of emphatic *does / doesn't* in affirmative sentences, and occurrences of these auxiliaries in both questions and VP ellipsis structures. The results for the input in the two children's files are provided in (i):

(i)

	Adam	*Eve*
Total number of adult utterances:	20,862	10,020
Total number of adult *does*:	408 (1.95%)	58 (0.57%)
Total number of adult *doesn't*:	226 (1.08%)	70 (0.69%%)
Total number of adult *does/doesn't*:	634 (3.03%)	128 (1.27%)

elements raise in the syntax in the adult grammar of English. The hypothesis is problematic for main verbs, however. Main verbs cannot raise to check the uninterpretable features that are generated in the inflection node, because the English setting of the Verb Raising Parameter requires verbs to stay inside the verb phrase. The Verb Raising Parameter is situated high in the parameter hierarchy, so, by hypothesis, it will already have been set at the point that children are considering the Inflection Parameter (cf. Wexler 1998). Thus the child confronts a dilemma. For main verbs, there is a conflict between the featural setting of the Inflection Parameter and the Verb Raising Parameter, which prevents main verbs from raising out of the VP.

How can the uninterpretable features in the inflection category be satisfied when the main verb cannot raise? We propose that, as children strive for a solution that can be implemented in English, their utterances contain "misplaced" morphology. This yields non-adult utterances such as: *He –s fit in there*. More specifically, the proposal is that children project *–s* as a phonologically weak auxiliary verb that raises to inflection to check off its uninterpretable features. The weak *–s* auxiliary verb, like *be* and *have*, raises out of the verb phrase to check off the uninterpretable features in INFL. As a weak clitic-like element, the *–s* auxiliary leans on its neighbors for support, but it need not attach to a verbal host. So, we take utterances like *He –s fit in there* as evidence that children have mis-set the Inflection Parameter and erroneously hypothesized featural inflection for English.[13]

The featural setting of the Inflection Parameter can be combined with either setting of the Negation Parameter. But since either setting of the Negation Parameter yields the same surface word order when combined with featural inflection, it is difficult to identify which setting of the Negation Parameter the child has selected. Following usual grammatical practice, if negation is a specifier, the auxiliary verb, here the weak *–s* auxiliary, can raise to INFL, resulting in examples like *He –s not fit in there.*[14] If negation is generated in the head of the phrase, the auxiliary verbs *have* and *be* are permitted to raise past negation (cf. Chomsky 1991), so the weak *–s* auxiliary is also expected to raise higher than negation — again, this would generate sentences like *He –s not fit in there*. Finally, if a child takes *don't* to be an unanalyzed negative chunk in head position, then utterances like *He –s don't fit in there* would also be expected. In Figure 6, we summarize the range of possible child utterances, both affirmative and negative, that are consistent with featural inflection.

13 Children were found to combine the stray morpheme with a range of NP types, including names and quantificational NPs. Since such NPs cannot be pluralized, the utterances like *Everybody –s fit* and *John –s fit* are evidence that children's non-adult forms were not plural marking, but some form of *–s* morpheme that was associated with inflection (cf. section 9.3).

14 In the syntactic literature, it is assumed that auxiliary verbs *have* and *be* can raise past the negative head without movement being blocked. It has been suggested this can happen because these verbs (or at least *be*) is semantically transparent. Although various accounts can be offered, it is basically a stipulation. Recall that when it comes to lowering of the affix over negation, the opposite assumption is made — that the negative head blocks movement.

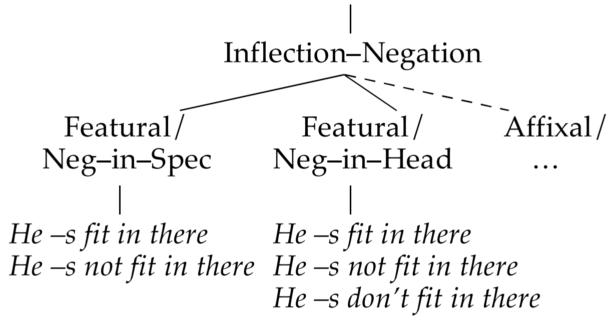


Figure 6: The utterances generated by the featural setting of the Inflection Parameter, with and without negation

Whether the child produces an affirmative or a negative sentence, the orphan *–s* morphology is a tell-tale sign that the child has a featural setting of the Inflection Parameter. The disappearance of the misplaced morphology, therefore, serves to flag the fact that such a child has switched from the featural to the affixal value of the parameter, the correct setting for adult English.

If the child's initial value of the Inflection Parameter is affixal, the child faces an even more arduous journey to the adult grammar. This is counterintuitive, since the affixal value is the target parameter setting for English. Furthermore, the hypothesis that English selects the affixal value of the Inflection Parameter works fine for affirmative sentences; the affix is simply lowered onto main verbs, resulting in adult-like utterances such as *He fits here*. The affixal value of the Inflection Parameter will prove problematic in negative sentences, however, if a child has not yet acquired *do*-support.[15] During the period before *do*-support is acquired, if the negative words *no* or *not* are taken to be positioned in the head of a phrase, then the affix is blocked from lowering past negation onto a verb inside the VP without violating the head movement constraint (Travis 1984, Chomsky & Lasnik 1993). Therefore, the child has no way to produce negative sentences without violating UG, despite having the correct adult setting of the parameter.

Children who find themselves in this quagmire could proceed in a number of ways. The most efficient route to the adult grammar would be to retain the affixal setting of the Inflection Parameter, and to reconcile this with positive input that contains the verbal element *does*.[16] The observation that *do* supports the *–s* affix in questions (*Does he fit in there?*), in sentences with VP ellipsis (*Yes, he does.)*, and in negative sentences (*He doesn't fit in there.*) would propel the child directly to the target grammar. Children who are unable to adopt *does* immediately into their lexicon, however, face a predicament. One way out for

---

15 Auxiliary verbs and modals are exceptions, of course, but children treating all verbs in a uniform way could presumably also lower an affix onto an auxiliary verb or a modal, without causing the derivation to crash. Our data set does not show evidence of this; there are no examples like *He cans fit here*, or *He be–s fit here*. Somehow, children must figure out that modals and auxiliaries behave differently early on.

16 We are assuming, for the moment, that the learner is not using a statistical mechanism to set the parameter.

them would be to produce utterances with no verbal morphology at all, such as *He no/not fit in there*.[17] In this case, such "optional infinitive" utterances (OI in Figure 7 below) can be seen as a last resort repair option taken by children who haven't acquired the form *doesn't*. An alternative strategy would be to try a different analysis, with negation as a specifier, on par with the negative adverb *never*. If this route is taken, the verbal *–s* affix is free to lower over *not*, yielding utterances like *He not fits in there* (cf. *He never fits in there*.).

The production types that result if children choose the affixal setting of the Inflection Parameter are summarized in Figure 7.

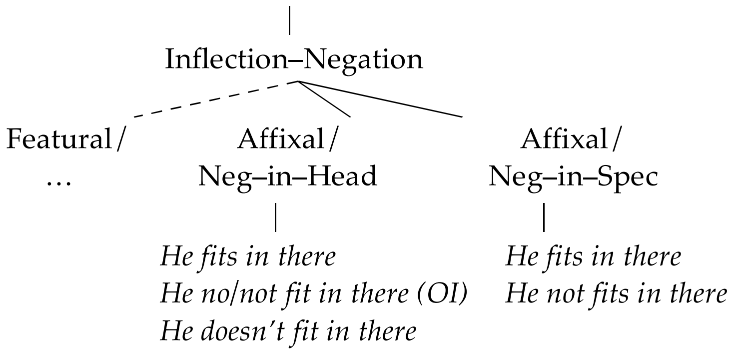


Figure 7: The utterances generated by the affixal setting of the Inflection Parameter with and without negation

It has been claimed in the acquisition literature that utterances such as *He not fits* would flout the head movement constraint (Harris & Wexler 1996) and the few cases found in the CHILDES database have been deemed to be performance errors. More recently, a head turning preference study with 19-month-old children by Soderstrom *et al.* (2002) found children prefer sentences like *At the bakery, a team not bakes bread* over the comparable optional infinitive version *At the bakery, a team not bake bread*. The child data we have collected are compatible with this result; we have observed many examples of *not* with an inflected verb. Assuming that children are treating negation as a specifier, the head movement constraint is not violated.

With these predictions in hand, the next section outlines the details about the child subjects, methodology and the empirical data gathered in studies of young children's acquisition of inflection and negation.

9. Acquisition of Inflection and Negation

The details of the longitudinal study conducted to examine children's development of inflection and negation are reviewed below.

---

17 Another option would be to sidestep violating the head movement constraint by placing negation outside the sentence, with utterances like *No he fit here*. Only one subject in our study, Georgia, used this option, but it surfaced before she was producing full sentences.

9.1. *Child Subjects*

The four children visited the language acquisition laboratory starting at about age 2, and they continued to visit the lab every two weeks for roughly a year, at which point the verbal morphology of the children was close to adult-like. The number of sessions and the duration of participation for each of the children in the study is given in Table 2.

Subject	*Age at beginning of study*	*Age at end of study*	*Number of sessions*
Caitlyn	1;09.04	2;08.29	18
Kristen	2;00.12	3;00.08	18
Georgia	1;10.23	2;08.20	19
Curtis	2;01.09	3;08.03	31

Table 2: Participants' ages and duration of participation in the study

9.2. *Methodology and Data*

The child data were collected using elicited production techniques in addition to data from spontaneous speech (cf. Crain & Thornton 1998). Because of the experimental component of the study, our data differ from that reported in much of the literature on 2-year-old English-speaking children's morphosyntax. Most developmental theories of early child language are based on naturalistic data obtained using transcriptions contained on the CHILDES database (MacWhinney 2000). Naturalistic data are insufficient for a detailed study of inflection in English-speaking children, however, because toddlers' play is often restricted to 'here and now' situations, comprised of talk between 'you and me'. As a consequence, few utterances with 3^{rd} person singular subject noun phrases are attested. Since only verbs corresponding to 3^{rd} person subject expressions bear agreement in English, transcript data is unlikely to contain sufficient data with 3^{rd} person subjects to reach firm conclusions about early acquisition of inflection. To rectify this problem, elicited production probes were used to boost the number of utterances with 3^{rd} person subjects and utterances with negation. These techniques enabled us to increase the amount of relevant 3^{rd} person data gathered for any one session, and therefore to follow children's development of verbal morphology more closely than would have been possible using only children's spontaneous speech.

The experimental protocols also revealed some types of utterances that have not been reported with any regularity before. In particular, sentences with misplaced morphology like *He –s fit in there*, and ones with medial negation like *He not fits in there* have previously been mentioned only in passing, and have usually been interpreted as performance errors because of their paucity. Our experimental techniques encouraged children to talk when they might otherwise have remained silent. This is particularly true of negative utterances, which are sparse in the naturalistic production represented in the CHILDES database. For example, Harris & Wexler (1996) searched the transcripts of 10 children who ranged in age ranged from 1;6 to 4;1, and found only 52 sentences with 3rd person subjects in structures that contained *no* or *not* and a main verb (cf. Harris & Wexler, Table 5:16). Our study evoked about 500 comparable structures from our four subjects over a considerably shorter period.

The elicited production procedures were straightforward. In order to elicit 3rd person subjects, the experimental workspace simply incorporated a third person, in addition to the child and the experimenter. The third person was usually a puppet, and children were asked questions about the puppet using yes/no questions (*Does the cat like milk?*). Children were also asked questions about other family members' likes and dislikes (*Does you daddy like milk?*). Procedures to evoke negative sentences included a range of games. One game focused on where various objects fit. For example, a puppet might try to complete a puzzle, but would end up putting pieces in the wrong place. The child was encouraged to correct the puppet (*It not goes there!*). Or, the child subject was assigned the task of testing groups of objects, to see if they float, or squeak, of would stick to a magnetic board, and so forth (*This one squeaks. This one not squeaks.*). These elicitation procedures resulted in a more robust set of data for each child that is available for the children whose data is housed in CHILDES.

The details of the longitudinal data collected for the four children are summarized in the tables below. Table 3 gives the details of the affirmative sentences. Altogether, the four children produced 2,044 affirmative sentences with 3rd person subjects. Of these, 1,381 contained verbs marked with 3rd person agreement morphology, and 663 had omissions of morphology and could therefore be considered to be root infinitives. The focus of this paper is on the 1381 utterances in which 3rd person morphology was expressed, as it is these sentences that are informative about the value children have adopted for the relevant parameters.

Type of affirmative utterance		*Type of 3rd person singular morphology*	
Utterances with 3rd person morphology	1,381 (68%)	Misplaced morphology Adult-like morphology	201 (15%) 1,180 (85%)
Utterances with omissions	663 (32%)		
Total utterances:	2,044		

Table 3: Total number of affirmative utterances produced by subjects (N=4)

The details of children's negative sentences are summarized in Table 4. The children produced a total of 506 negative utterances with 3rd person singular subjects. Of these, 322 were finite sentences. It is these finite sentences that are crucial for tracking children's settings of the Negation Parameter.

Type of negative utterance		*Type of 3rd person singular morphology*	
Utterances with 3rd person morphology	322 (64%)	Misplaced morphology	30 (9%)
		Medial negation	98 (30%)
		Adult *do*-support	194 (61%)
Utterances with omissions	184 (36%)		
Total utterances:	506		

Table 4: Total number of negative utterances produced by subjects (N=4)

Summarizing the data from Tables 3 and 4, the four children produced a total of 2,550 sentences with 3rd person singular subjects. The 1,703 of the 2,550 that bear some kind of agreement morphology form the basis for our evaluation of the alternative parameter-setting models. Before we proceed with the evaluation, a word is in order about the exceptional types of utterance that children produced.

9.3. Novel Utterances

We mentioned one kind of novel utterance earlier; for example *He –s fit in there.* These utterances with "misplaced morphology" have not been reported in previous literature as a grammatical option for children. The phenomenon was observed by Stromswold (1990) in the transcripts of Adam (in the CHILDES database), although Adam's use was restricted to the pronoun *it* and did not appear with other 3rd person subjects.[18] Instances of misplaced inflection have also been reported by Foley *et al.* (2003), in a study using an elicited imitation methodology. In the child data we obtained, misplaced inflection was not limited to experimental situations. However, elicited production techniques evoked sufficient numbers of examples to demonstrate that the misplaced inflection is clearly a grammatical option for some children. Since these novel utterances are not well documented, we lay out here why we consider them to be a grammatical option for some children.

We can begin by noting that if these utterances were an artifact of our experimental techniques they should appear for all children. This is not the case. As will become clear, only 3 of the 4 children produce such utterances. Furthermore, if the misplaced morphology were artifactual, one would not expect it to have a specific syntactic distribution. But it does. This misplaced morpheme appeared only with 3rd person subjects in present tense contexts, and not in sentences with a modal or auxiliary verb. So no child produced a sentence like *He cans fit*, or *He –s is eating*, for example.

An alternative explanation is that the misplaced morphology is, in fact, a plural morpheme for the subject noun phrase (as in *The cats fit in there.*). All sessions with children were videotaped, however, and it is clear that children

18 In our view, it is likely that misplaced morphology has often been mis-transcribed as a plural marker.

produced such utterances when they were referring to single objects. One might argue that even though the child sees only one cat in the workspace, the utterance is about cats in general. If this is the case, one could ask why this plural morpheme occurred only in 3rd person present contexts. Furthermore, the children's utterances demonstrate that the misplaced *–s* shows up with a range of NP types (i.e. common nouns, names, singular pronouns, and quantificational nouns like *everybody*). Since names, singular pronouns and quantificational NPs cannot appear with a plural *–s* morpheme, analysis of the extra morphology as a plural morpheme is not tenable. Data from the child subjects cements the point. For Kristen, around 65% of her utterances with misplaced inflection were with singular pronouns *he*, *she*, *it*. For Georgia, who produced well over a hundred instances of misplaced morphology, 40% of the cases were with proper names, including 25 instances where she uses her own name (as in utterances like *Georgia –s like it.*). It is highly unlikely that children would use the plural morpheme with their own name as some kind of generic use of the plural. For all of these reasons, we conclude that the stranded *–s* is, indeed, 3rd person agreement morphology.

Utterances with medial negation, such as *He not fits in ther*e, were also observed, and for the first time, enough of these utterances were observed to conclude that they are consistent with children's grammars. In a previous search of the CHILDES database, Harris & Wexler (1996) found few such occurrences, which invited the conclusion that they were performance errors. The present data demonstrate clearly that, at least for some children, medial negation utterances are a grammatical option, on our analysis, representing a mis-set parameter.

10. Evaluation of the Models

This section discusses the three main criteria that distinguish between the Hierarchical Acquisition model and the Variational model: (i) initial value, (ii) trajectory, and (iii) conformity. In each case, the longitudinal data from the four child subjects will be used to assess how well the accounts stand up to the empirical findings. The Inflection Parameter is discussed first, and then the Negation Parameter.

10.1. The Inflection Parameter

Triggering models, including the Hierarchical Acquisition model, anticipate that children will consistently apply one parameter value unless parameter-resetting is required. As we noted, however, if the type of sentence that is indicative of one or other parameter value is optional, then its use may not reach the 90% correct usage criterion of grammatical knowledge. We witnessed this in children's use of null subjects, where children used both null subjects and lexical subjects, until pronominal subjects replaced the null subjects. The child production data relevant to the Inflection Parameter likewise shows two forms. On our analysis, this is because adult-like sentences such as *He fits* also appear at the stage when children produce non-adult utterances such as *He –s fit.*[19] For children who have the featural value of the Inflection Parameter, there may be more than one way to realize the value.[20] Since there is no way to tell what analysis children are assigning to what look like adult utterances (*He fits.*) in the early stages, we have omitted adult-like utterances in the counts of featural inflection. But it should be kept in mind that this necessarily reduces the proportion of utterances representing the featural value of the Inflection Parameter in the graphics we present.

10.1.1. Trajectory

Children's trajectories for the Inflection Parameter are summarized in Table 5. As can be seen, the children chose one or the other value of the parameter as their starting point; two children selected the featural value of the parameter, and two children selected the affixal value. There was no child for whom both values seemed to be competing in the earliest stages of acquisition. This finding is not anticipated on the Variational model. Of the two children who selected the non-adult featural value, one child (Georgia) also exhibited abrupt parameter-resetting, eliminating the misplaced morphology and switching to adult-like utterances. The other child (Kristen) exhibited a smaller drop in use of non-target utterances, but the drop itself was nevertheless quite precipitous.

We consider next the two children who selected the adult affixal value of the Inflection Parameter. One child (Caitlyn) rapidly converged on the adult grammar, while the other child (Curtis) meandered, taking well over a year longer to acquire the adult value for the parameter. This child initially set the parameter to the affixal value, then reversed his setting to the non-target featural value, and then reversed the setting yet again, finally deciding that English does, after all, take the affixal value of the parameter.

Subject	*Initial value*	*Trajectory*	*Parameter stable*
Georgia	featural	abrupt	2;4
Curtis	affixal	gradual?	3;3
Kristen	featural?	abrupt?	2;7
Caitlyn	affixal	none	2;0

Table 5: Summary of initial value and trajectory for Inflection Parameter

19 When children hypothesize the featural option, the *–s* morphology is a clitic-like element. Cross-linguistically, clitics are much freer than affixes in where they may be positioned, and they may lean to the left or right in the sound stream. Keeping these properties in mind, in the featural stage, children could have two possible realizations of the clitic; one preceding the main verb in *He –s fit*, and one after the verb in *He fit –s*. This behavior would be similar to Polish person–number agreement marker *śmy*, which can appear on any constituent preceding the verb but to no element following it (Franks & Bański 1998, Witkoś 1998).

20 Notice that the optionality does not represent two different parameter settings.

The detailed trajectories for the individual children are shown in a series of graphs, which all show the decline of utterances with misplaced morphology as a percentage of all affirmative and negative utterances that contain an inflected verb. Since the adult utterances (*He fits.*) do not, in principle, distinguish between featural and affixal morphology, they cannot be decisively used to show the introduction of affixal morphology. For this reason, the increase in adult-like utterances is not represented on the graphs.

The data for Georgia are depicted in Figure 8.

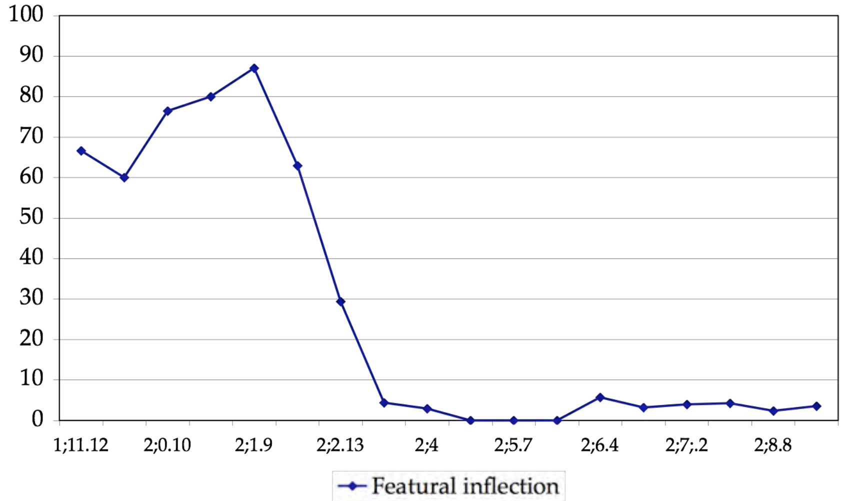


Figure 8: Proportion of featural inflection across sessions for Georgia

Altogether, Georgia produced 591 affirmative and negative finite sentences, of which 144 have misplaced morphology. The data are graphed from the second session, since Georgia produced only one verb in the first session at 1;10.23.[21] In her session at 1;11.12, Georgia used featural inflection (*He –s fit in there.*) in 67% of her finite utterances. This then increased to 87% at 2;1.9, despite the absence of evidence in the input corresponding to this setting of the parameter. By 2;3.16, just 2 months later, Georgia's use of featural inflection had declined to 4% of utterances with an inflected verb, an 83% decline. Once the featural option dropped out, it did not constitute more than 5-6% of Georgia's productions in any subsequent session. This kind of trajectory is not the gradual curve that is associated with statistical learning, and seems, instead, to be indicative of categorical change, as expected by the Hierarchical Acquisition model.

The graph in Figure 9 shows data from Curtis. These data reveal a more chaotic path than the one taken by Georgia, with non-target utterances extending over a longer period of time. During the time period indicated in the graph, Curtis produced 505 affirmative and negative finite sentences with 3rd person subjects, 37 of which have misplaced morphology.

[21] In all of the children's graphs for featural inflection, the graph starts from the session in which 5 or more utterances with finite verbs were used.

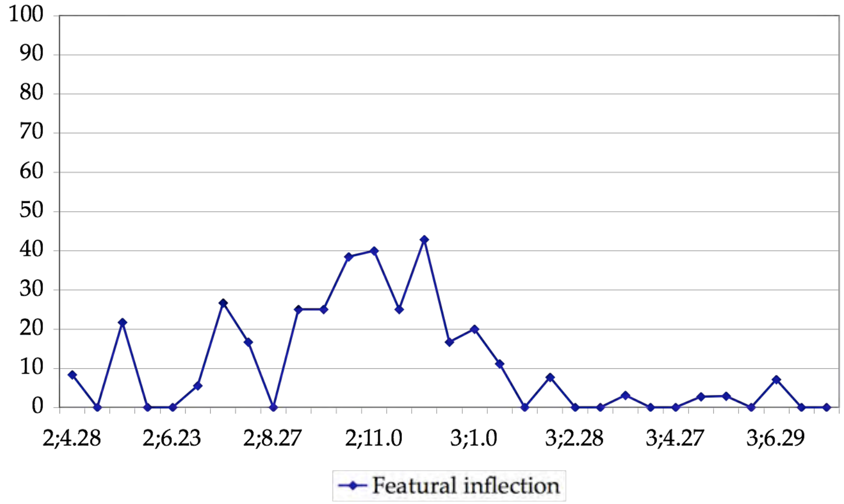


Figure 9: Proportion of featural inflection across sessions for Curtis

Initially, Curtis appears to have adopted the affixal value for the inflection parameter. Then, he appears to switch to the featural value, only to switch back again. In the first session with productive use of finite sentences, featural inflection was exhibited less than 10% of the time (if misplaced inflection is used as the yardstick). Featural inflection then rose gradually, taking several months to reach asymptote. Since there is no evidence for featural inflection in the input, this learning curve is not easily reconciled with the Variational model. Featural inflection reached 43% use at 2;11.28. However, within three months, Curtis's featural inflection was eliminated, with the adult parameter setting becoming stable at about 3;3. Thus, although the trajectory is prolonged, when the parameter is set conclusively, the change takes place within 3 months. The timeline suggests that Curtis was able to ignore the distributional data in the input until almost 3 years of age, at which time parametric change was instigated.

The trajectory for Kristen is depicted in Figure 10.

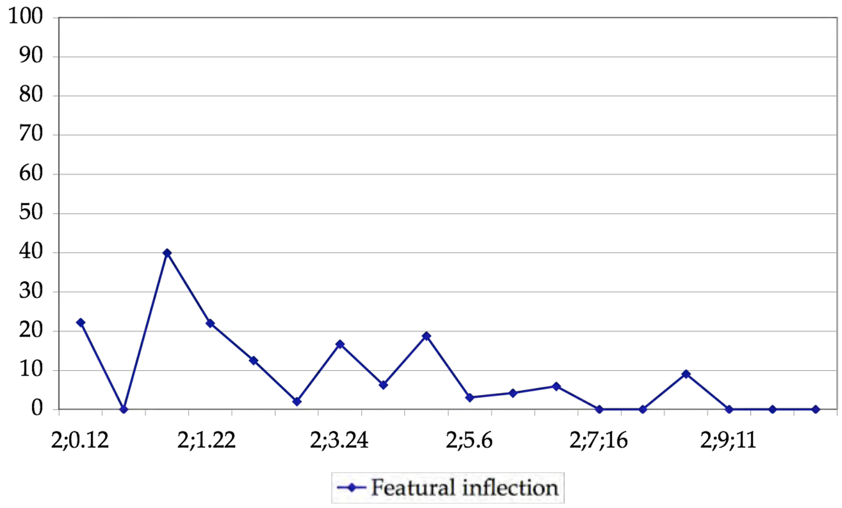


Figure 10: Proportion of featural inflection across sessions for Kristen

Kristen's use of featural inflection is graphed beginning with the first session at 2;0.12. In all of the sessions, she produced a total of 409 finite sentences with 3rd person subjects, 44 of which have misplaced morphology. However, all indications are that the onset of her production of inflection had already begun before the first session at our language acquisition laboratory. During the first session, Kristen used featural inflection in 22% of her finite sentences. This rose to 40%, but never exceeded this percentage. Although Kristen used the featural option of the Inflection Parameter at most 40% of the time, as seen at age 2;1.6, the percentage had dropped to 2% by 2;3.7, i.e. in just two months. After that, its use remained below 10% in all but two of the sessions. Again, there is a fairly sharp change.

The graph in Figure 11 illustrates the pattern of development of the fourth child, Caitlyn.

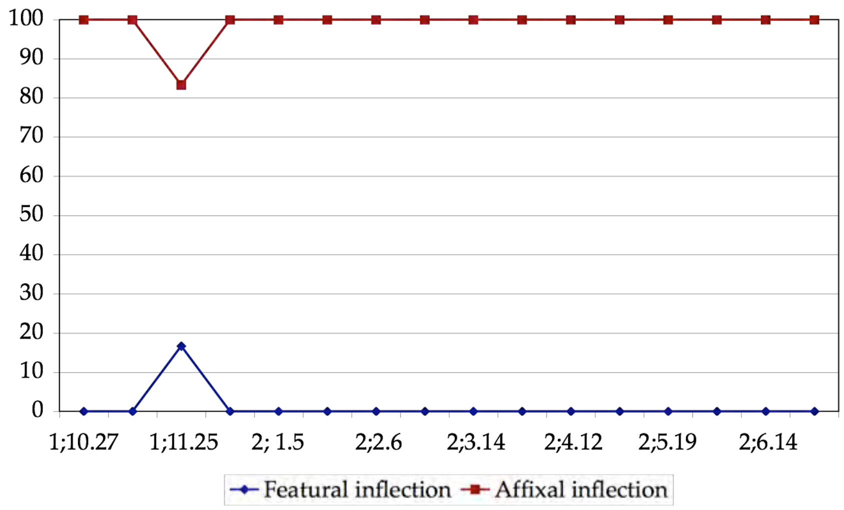


Figure 11: Proportion of featural inflection across sessions for Caitlyn

During this time, Caitlyn produced 198 finite sentences with a 3rd person subject. The small peak in the graph at 1;11.25 represents the 5 examples Caitlyn produced that can be interpreted as evidence of the featural value of the Inflection Parameter.[22] Thus Caitlyn had the adult value of the Inflection Parameter from the start, and no change in the parameter value was observed. It is our interpretation of Caitlyn's data, then, that the Inflection Parameter was initially set to the affixal value, hence Caitlyn's adult-like productions such as *He fits* are taken to unambiguously reflect the affixal parameter value, and for this reason, they are included in Caitlyn's graph.

[22] Caitlyn's graph shows her data from the third session. The first two sessions are not graphed as both sessions contained only two finite sentences.

10.1.2. Conformity

Another distinguishing criterion is conformity in acquisition across children. This is expected on the Variational model, but not on the Hierarchical Acquisition model. The profiles of the four children clearly do not display conformity. The children take different paths to the adult value of the Inflection Parameter, and achieve the adult value of the parameter at different rates. Caitlyn initially selected the affixal value and changed abruptly to become adult-like by 2;0. The children who initially selected the featural value (Kristen and Georgia) switched within a few months to the affixal value. Moreover, one child, Curtis, settled on the affixal value an entire year later than other children did. This child did not change to the correct value until he was over 3 years of age. This was not due to gradual learning, however. In fact, Curtis manifested a "pendulum" learning curve, switching from the affixal value to the featural value of the Inflection Parameter, and then back again to the affixal value. To sum up, different children begin with different start values, take different paths, and reach the "final state" at different rates and at different times. Conformity is not characteristic of children's behavior.

10.2. The Negation Parameter

The Negation Parameter gives the learner two options in assigning a position to negation in the phrasal structure of sentences. The predictions of the Hierarchical Acquisition model and the Variational model will be discussed for this parameter using the same three criteria: (i) initial value, (ii) trajectory, and (iii) conformity.

10.2.1. Trajectory

The data indicate that, in the initial stages, children select one parameter value or the other, but not both values of the Negation Parameter. Georgia hypothesizes the adult value with negation residing in head position, whereas the other 3 children begin with negation located in specifier position. Since Georgia begins with the adult value, the trajectory of parameter-setting for Georgia is essentially flat, although she needs to acquire the lexical item *doesn't*. The productions of the 3 other children, who initially adopt the non-target value, exhibited an abrupt change in values, as anticipated on the Hierarchical Acquisition model. Moreover, the precipitous changes manifested by different children were initiated and completed at different times. There was no indication that the statistical distribution of structures or lexical items in the input was responsible for the trajectories of any of the children. If the assumption of uniformity of input is correct, then children's data corresponding to the Negation Parameter do not provide support for the Variational model. The data for the 4 children are summarized in Table 6.

Subject	*Initial value*	*Trajectory*	*Parameter stable*
Georgia	Neg–in–Head	none	2;6
Curtis	Neg–in–Spec	abrupt	3;7
Kristen	Neg–in–Spec?	abrupt?	2;7
Caitlyn	Neg–in–Spec	abrupt	2;0

Table 6: Summary of the trajectory and initial value for Negation Parameter

The trajectory for each child is shown below in a series of graphs.[23] Recall that the same value of the Negation Parameter may have different surface manifestations, at different points in time, depending on the child's current hypothesis about the value of the Inflection Parameter. In particular, a child who has hypothesized the Neg–in–Spec value of the Negation Parameter might produce sentences of the form *He –s not fit in there* at the stage at which the Inflection Parameter is set to the featural value, but later, when the parameter is switched to the affixal value, the same (Neg–in–Spec) value of the Negation Parameter would result in sentences of the form *He not fits there*. For the relevant children, two graphs illustrate the course of development. The first graph shows the varying surface manifestations of the hypothesized parameter value; the second graph collapses these variations of surface forms, to more clearly display the development of the Neg–in–Head value versus the Neg–in–Spec value of the Negation Parameter.

The data for Georgia are shown in Figures 12 and 13.

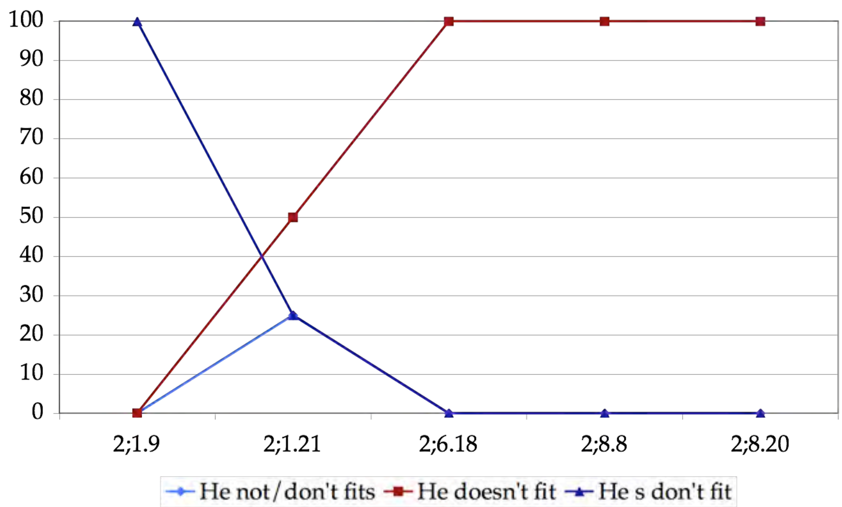


Figure 12: Negative forms used by Georgia

[23] The graphs start with the session in which the child first produced two or more finite sentences. After that, the data for any session containing a finite sentence and negation was included in the graph.

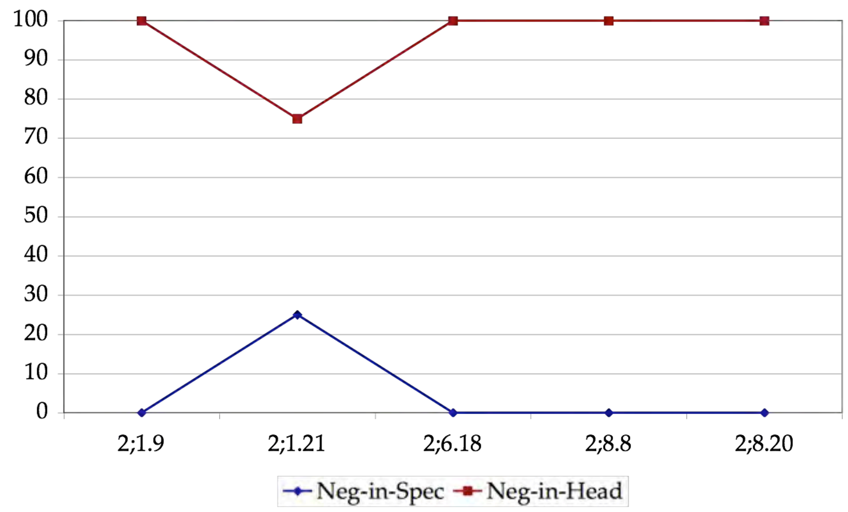


Figure 13: The trajectory of the two negation values in Georgia's data

There were only 5 sessions in which Georgia produced negative sentences, so the data in the graphs is limited to these 5 sessions in which 29 instances of negation were produced with a 3rd person subject.[24] Although Georgia's data are limited, it is clear that the adult Neg–in–Head value of the Negation Parameter was her initial hypothesis. This child did not use the item *not* at all. The few negative utterances Georgia did produce incorporated *do* as a host for the head *n't*, in sentences of the form *He –s don't fit*. These examples are interpreted as a reflex of the Neg–in–Head value of the parameter, combined with the featural value of the Inflection Parameter. Since Georgia did not mis-set the Negation Parameter, the graph does not reveal one value of the Negation Parameter being replaced by another. However, Georgia does mis-set the Inflection Parameter. The effect of this is revealed in the sentences she produced that illustrate the Neg–in–Head value of the Negation Parameter across time. The different negative forms used by Georgia are given in Figure 12. Before the featural value of the Inflection Parameter is eliminated, Georgia's negative sentences are of the form *He –s don't fit*, but once Georgia acquires the affixal value of the Inflection Parameter, at around 2;6, her negative sentences contain *doesn't*. Since these two utterance types are both evidence of the Neg–in–Head value of the negation, it is reasonable to collapse them, which results in Figure 13. This figure illustrates that Georgia maintains a constant value for the Negation Parameter. In the session at 2;1.21, one of Georgia's 4 utterances shows medial negation, with negation in the specifier position (i.e. *It don't squeaks.*[25]), hence the small peak in the graph.

---

[24] The data set is small because Georgia resisted producing negative sentences until she acquired *do*-support. When elicited production techniques were used to probe negation, her preferred strategy was to use covertly negative elements like *only* or *just* to answer questions. For example, in response to a question like *Does your daddy drink milk?*, rather than answering *Daddy doesn't like milk*, she would say *Only Georgia drinks milk*.

[25] This is considered to be medial negation if *don't* is taken to be a chunk that expresses negation. This is the case for Georgia This child did not use *not* at all. She did use *no* as a form of sentence external negation in the earliest recordings, however.

The data for Curtis are summarized in Figure 14.

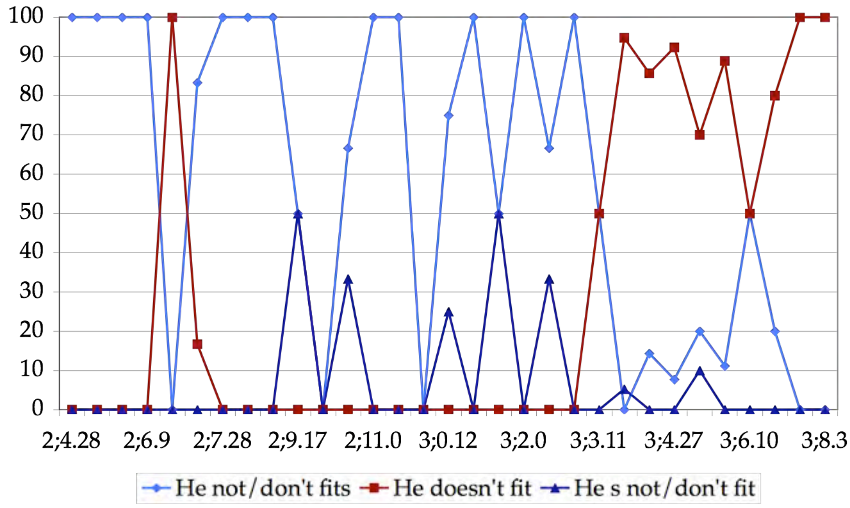


Figure 14: Negative forms used by Curtis

The profile appears chaotic because Curtis produces a variety of negative forms. Altogether, he produced 181 negative utterances with a 3rd person subject in a finite sentence. However, as with Georgia, the fluctuation for Curtis can be attributed to the value of the Inflection Parameter, which affects the form of Curtis's negative sentences. During the first few sessions, Curtis had a single initial value for the Negation Parameter; 100% of Curtis's productions exemplify the Neg–in–Spec value. Optionality in forms appears only later in the negative sentences produced by Curtis. This is shown in Figure 14. However, the optionality exhibited by Curtis is evidence of his change in values for the Inflection Parameter, and does not involve the Negation Parameter. Curtis starts with the affixal value of the Inflection Parameter, but then switches to the featural value (perhaps in order to permit him express negation). At that point, Curtis produced utterances with misplaced inflection, as in sentences of the form *He –s not fit in there*. Later, Curtis reverted to the affixal value of the Inflection Parameter and finally, *doesn't* appears.

In presenting the developmental trajectory of Curtis, we have combined the later negative utterances with featural inflection (*He –s not fit in there.*) and the earlier medial negation utterances (*He not fits in there.*) in Figure 15 below because both forms reflect a constant Neg–in–Spec value for the Negation Parameter. Once these alternative forms are combined, the chaotic peaks in the earlier graphic representation flatten out considerably. As Figure 15 shows, Curtis uses the Neg–in–Spec value of the Negation Parameter until about age 3;3.[26]

[26] One might try to challenge the decision to take negative items with misplaced morphology, like *He –s not fit in there* as reflecting Neg–in–Spec since such utterances can be the product of either setting of the Negation Parameter. If the utterances with misplaced morphology were taken to reflect Neg–in–Head, then Curtis's data would show more gradual acquisition of the Neg–in–Head parameter value, although the medial negation would still drop

Apparently Curtis ignored the relevant input for many months. Finally, at about 3 years of age, Curtis's grammar abruptly changed. The change culminated at about 3;3. In the session at 3;2.28, Curtis produced 12 examples with medial negation. A month later, when he was 3;3.30, medial negation had disappeared, and Curtis produced 18 cases of *doesn't* in a single session. Again, the evidence from Curtis is difficult to reconcile with the Variational model, first, because the positive input appears to have no impact on Curtis's productions for many months, and second, when change does take place, it is swift.

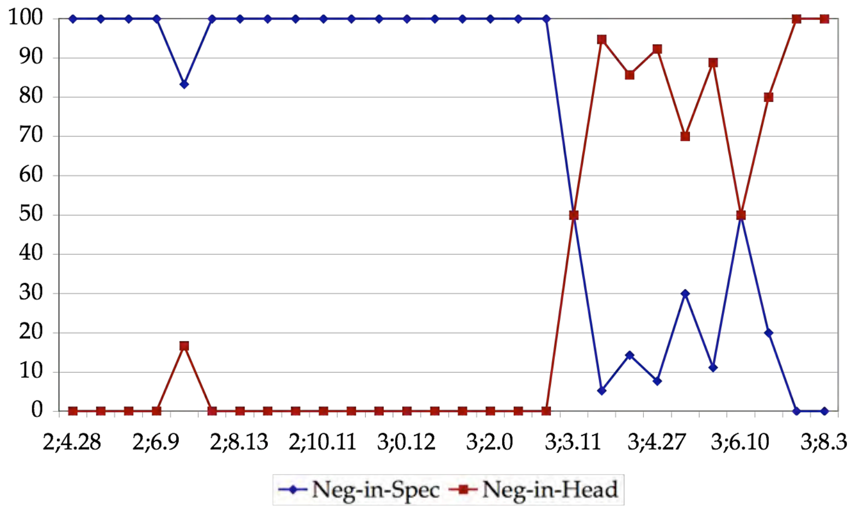


Figure 15: The trajectory of the two negation values in Curtis's data

Kristen used a variety of negative forms, as shown in Figure 16.

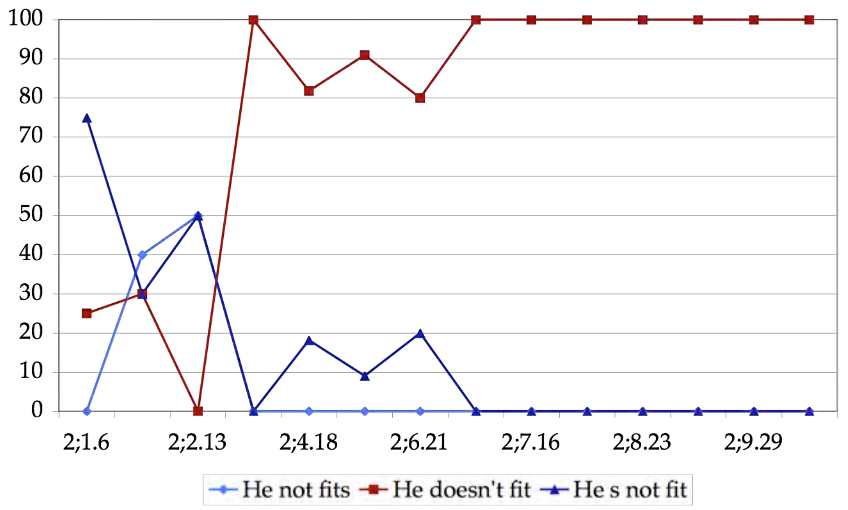


Figure 16: Negative forms produced by Kristen

abruptly. Curtis would also be allowing the item *not* to be sometimes positioned in the head and sometimes in the spec, rather than reserving different items for the different positions.

Altogether, Kristen produced 69 finite negative utterances with a 3rd person subject. Kristen's data do not reflect the onset of her production, so it is not possible to be certain of her initial value of the Negation Parameter. Judging from Figure 16, it seems most likely that the initial value of the parameter was Neg–in–Spec. In the first session recorded in the graph, there were four utterances with negation in a finite sentence; 3 with misplaced morphology and one adult-like example with *doesn't*. As with Curtis, the optional forms are more likely to be the product of the Inflection Parameter, rather than the Negation Parameter. At first, Kristen appears to use both medial negation (i.e. the affixal value of the Inflection Parameter) and misplaced morphology (the featural value).

If the medial negation utterances and the negative utterances with misplaced morphology are collapsed, as we did with Curtis's data, and both utterance types are taken to represent the Neg–in–Spec option of the Negation Parameter, the pattern shown in Figure 17 emerges.

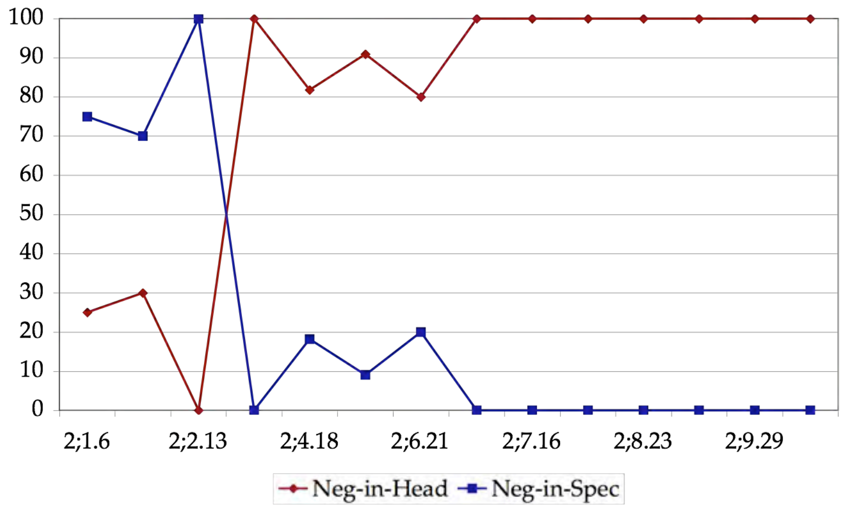


Figure 17: The trajectory for the two negation values in Kristen's data

There is sharp change between 2;2 and 2;3 as the Neg–in–Spec value of the parameter is switched to the Neg–in–Head value and utterances with *doesn't* begin to appear. Once the Neg–in–Spec value of the Negation Parameter has been replaced, utterances with *doesn't* suddenly appear; there are no examples with *doesn't* in the session at 2;2.13, but it is present in 9 of the 11 negative sentences produced by Kristen in the session at 2;4.18. The trajectory in Figure 17 above shows the precipitous change (here within 2 months) that is anticipated by the Hierarchical Acquisition model.

The trajectory for the negation pattern for Caitlyn is shown in Figure 18.

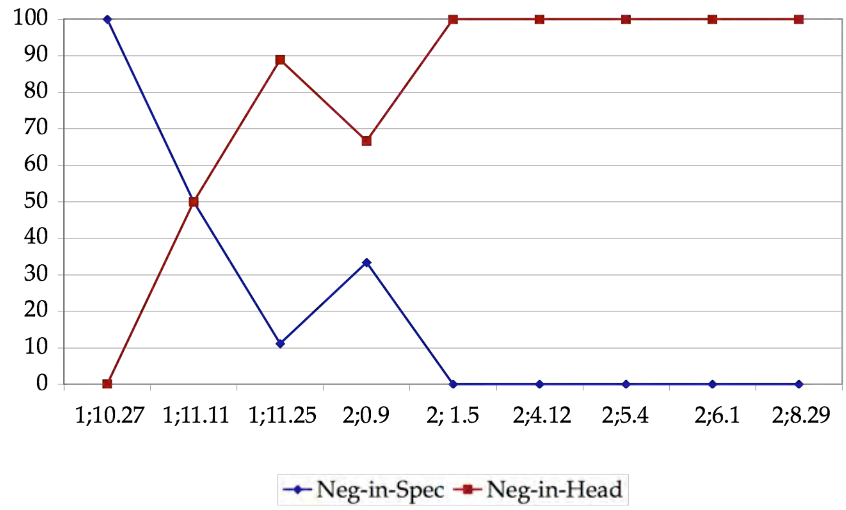


Figure 18: The trajectory for the two negation values in Caitlyn's data

Caitlyn produced 43 finite negative sentences with a 3rd person subject. Caitlyn initially selected the Neg–in–Spec value of the Negation Parameter, using only this option in the session at 1;10.27, when she produced 3 instances of medial negation. Grammatical change was initiated almost immediately, and medial negation rapidly disappeared, and was completely gone within 3 months. Recall that for the Inflection Parameter, Caitlyn hypothesized the affixal value of the Inflection Parameter from the start. While this would be an unfortunate choice before *do*-support is acquired, Caitlyn took heed of the input early, and quickly became adult-like, acquiring *doesn't* by age 2;1. Another interesting observation is that, at 1;11.11, Caitlyn produced 50% medial negation sentences and 50% adult-like sentences, with *do*-support. In fact, in the first half of the session, Caitlyn produced 6 utterances containing medial negation, and in the second half, she produced 6 adult-like utterances. In other words, the abrupt change from one parameter value to the other took place within a single session at the lab.

10.2.2. Conformity

It should be clear that the data do not fit neatly with the criterion of conformity. Three of the children initially mis-set the Negation Parameter, while one child started with the adult Neg–in–Head value of the parameter. The three children who had mis-set the parameter made abrupt changes in switching the parameter value to Neg–in–Head, but the change was initiated at different times for different children. For example, Caitlyn initiated change at about 2 years, while Curtis waited until about 3 and a half years of age. Thus it appears that the parameter-setting mechanism of different children does not respond in the same way to the presumably uniform statistical distribution of sentence structures in the positive input. One possibility to consider about the source of the timing differences is that children were delayed by the course they had taken in setting parameters that sit higher on the hierarchy.

11. Conclusion

Beginning with Saffran *et al.* (1996), the last decade has seen a series of research studies showing that children are endowed with a learning mechanism that is sufficiently powerful to assist them in word segmentation, and even in the detection of phrasal units (Saffran 2001, 2002). Yang (2002, 2004) has proposed that such learning mechanisms can also be paired with UG to assist the language learner in keeping tally of the input data necessary for setting parameters. Granting that learners employ a statistical learning mechanism for certain tasks, the empirical thrust of the present paper was to assess the claim that children make use of such a mechanism in setting parameters.

To address this question, we investigated children's acquisition of two parameters, to see whether the learning path in child language development assumed the gradual curve associated with statistical learning over time or, instead, if the path of language development resembled the sharp edges associated with setting and resetting parameters, in keeping with the triggering models. The empirical findings from our longitudinal study of four children's development of inflection and negation do not support the proposal that statistical learning is driving children's parameter-setting. Our empirical findings show, instead, that children initiate grammatical change at some point in time, and when change is initiated, it takes hold quickly, and is brought to closure within 3 months. The debate over what constitutes gradual learning and what constitutes triggering will no doubt continue in the literature. However, the observations made in this paper leave open the possibility that the mechanisms used to set parameters are specific to the language faculty, and do not consist of domain general statistical learning mechanisms, as Yang (2002, 2004) proposes. At this point, we do not fully understand the mechanisms that set grammatical change in motion, but they are apparently sensitive to the child's internal grammatical development, and do not directly reflect children's linguistic experience.

References

Armon-Lotem, Sharon, Stephen Crain & Spyridoula Varlokosta. 2004. Interface conditions in child language: Cross-linguistic studies on the nature of possession. *Language Acquisition* 12(3–4), 171-217.

Atkinson, Martin. 1992. *Children's Syntax: An Introduction to Principles and Parameters Theory* (Blackwell Textbooks in Linguistics 5). Oxford: Blackwell.

Baker, Mark. 1996. *The Polysynthesis Parameter* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Baker, Mark. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar.* New York: Basic Books.

Baker, Mark. 2005. Mapping the terrain of language acquisition. *Language Learning and Development* 1(1), 93-129.

Becker, Misha. 2000. The development of the copula in child English: The lightness of *be*. Los Angeles, CA: University of California dissertation.

Berwick, Robert & Amy Weinberg. 1984. *The Grammatical Basis of Linguistic Performance: Language Use and Acquisition* (Current Studies in Linguistics 11). Cambridge, MA: MIT Press.

Bobaljik, Jonathan David. 1994. What does adjacency do? *MIT Working Papers in Linguistics (MITWPL)* 22, 1-32.

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

Bloom, Lois. 1970. *Language Development: Form and Function in Emerging Grammars.* Cambridge, MA: MIT Press.

Bloom, Paul. 1990. Subjectless sentences in child language. *Linguistic Inquiry* 21(4), 491-504.

Borer, Hagit. 1984. *Parametric Syntax*. Dordrecht: Foris.

Borer, Hagit & Kenneth Wexler. 1987. The maturation of syntax. In Roeper & Williams (eds.), 123-172.

Borer, Hagit & Kenneth Wexler. 1992. Bi-unique relations and the maturation of grammatical principles. *Natural Language & Linguistic Theory* 10(2), 147-190.

Brown, Roger. 1973. *A First Language: The Early Stages*. Cambridge, MA: Harvard University Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use* (Convergence). New York: Praeger.

Chomsky, Noam. 1991. Some notes on economy of derivation and representation. In Robert Freidin (ed.), *Principles and Parameters in Comparative Grammar* (Current Studies in Linguistics 20), 417-454. Cambridge, MA: MIT Press. [Reprinted in Chomsky (1995), 129-166.]

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.

Chomsky, Noam. 2002. *On Nature and Language*. Cambridge: Cambridge University Press.

Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In Joachim Jacobs, Arnim von Stechow & Wolfgang Sternefeld (eds.) *Syntax: Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science 9/1), 506-569. Berlin: Mouton de Gruyter. [Reprinted in Chomsky (1995), 13-127.]

Clahsen, Harald. 1990. Constraints on parameter setting: A grammatical analysis of some acquisition stages in German child language. *Language Acquisition* 1(4), 361-391.

Crain, Stephen. 1991. Language acquisition in the absence of experience. *Behavioral and Brain Sciences* 14(4), 597-650.

Crain, Stephen, Wei Jia Ni & Laura Conway. 1994. Learning, parsing and modularity. In Charles Clifton, Lyn Frazier & Keith Rayner (eds.), *Perspectives on Sentence Processing*, 443-468. Hillsdale, NJ: Lawrence Erlbaum.

Crain, Stephen & Paul Pietroski. 2002. Why language acquisition is a snap. *The Linguistic Review* 19(1–2), 163-183.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics* (Language, Speech, and Communication). Cambridge, MA: MIT Press.

Fodor, Janet Dean. 1998. Unambiguous triggers. *Linguistic Inquiry* 29(1), 1-36.

Foley, Claire, Zelmira Núnez del Prado, Isabella Barbier & Barbara Lust. 2003. Knowledge of variable binding in VP-ellipsis: Language acquisition research and theory converge. *Syntax* 6(1), 52-83.

Franks, Steven & Piotr Bański. 1998. Approaches to "schizophrenic" Polish person agreement. *Formal Approaches to Slavic Linguistics (FASL)* 7, 123-144.

Gibson, Edward & Kenneth Wexler. 1994. Triggers. *Linguistic Inquiry* 25(3), 355-407.

Goldberg, Adele. 2003. Constructions: A new theoretical approach. *Trends in Cognitive Science* 7(5), 219-224.

Goro, Takuya. 2004. On the distribution of *to*-infinitives in early child English. *Boston University Conference on Child Language (BUCLD)* 28(1), 199-216.

Grinstead, John. 1994. Consequences of the maturation of number morphology in Spanish and Catalan. Los Angeles, CA: University of California MA thesis.

Guilfoyle, Eithne. 1984. The acquisition of tense and the emergence of lexical subjects in child grammars. *McGill Working Papers in Linguistics (McGWPL)* 2(1), 20-31.

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

Harris, Tony & Kenneth Wexler. 1996. The optional infinitive stage in child English: Evidence from negation. In Harald Clahsen (ed.), *Generative Perspectives on Language Acquisition: Empirical Findings, Theoretical Considerations and Crosslinguistic Comparisons* (Language Acquisition and Language Disorders 14), 1-42. Amsterdam: John Benjamins.

Hyams, Nina. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Hyams, Nina. 1987. The theory of parameters and syntactic development. In Roeper & Williams (eds.), 1-22.

Hyams, Nina. 1989. The null subject parameter in language acquisition. In Osvaldo Jaeggli & Ken Safir (eds.), *The Null Subject Parameter* (Studies in Natural Language and Linguistic Theory 15), 215-238. Dordrecht: Kluwer.

Hyams, Nina. 2007. Aspectual effects on interpretation in early grammar. *Language Acquisition* 14(3), 231-268.

Hyams, Nina & Teun Hoekstra. 1998. Aspects of root infinitives. *Lingua* 106(1-4), 81-112.

Hyams, Nina & Kenneth Wexler. 1993. On the grammatical basis of null subjects in child language. *Linguistic Inquiry* 24(3), 421-460.

Jaeggli, Osvaldo & Nina Hyams. 1988. Morphological uniformity and the setting of the null subject parameter. *North Eastern Linguistics Society* (*NELS)* 18, 238-253.

Lasnik, Howard 1981. Restricting the theory of transformations: A case study. In Norbert Hornstein & David Lightfoot (eds.), *Explanation in Linguistics: The Logical Problem of Language Acquisition*, 152-173. London: Longman.

Lasnik, Howard. 1995. Verbal morphology: *Syntactic Structures* meets the minimalist program. In Hector Campos & Paula Kempchinsky (eds.), *Evolution and Revolution in Linguistics: Essays in Honor of Carlos Otero*, 251-275. Washington, DC: Georgetown University Press. [Reprinted in Lasnik (1999), 97-119.]

Lasnik, Howard. 1999. *Minimalist Analysis* (Generative Syntax 3). Malden, MA: Blackwell.

Lasnik, Howard. 2001. A note on the EPP. *Linguistic Inquiry* 32(2), 356-362.

Lasnik, Howard & Stephen Crain. 1985. On pronominal reference. *Lingua* 65(1–2), 135-154.

Lebeaux, David. 1988. Language acquisition and the form of the grammar. Amherst, MA: University of Massachusetts dissertation.

Lebeaux, David. 1987. Comments on Hyams. In Roeper & Williams (eds.), 23-40.

Legate, Julie & Charles Yang. 2007. Morphosyntactic learning and the development of tense. *Language Acquisition* 14(3), 315-344.

Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution* (Blackwell-Maryland Distinguished Lecture Series on Language and Cognition 1). Malden, MA: Blackwell.

Lillo-Martin, Diane. 1986. Effects of the acquisition of morphology on syntactic parameter setting. *North Eastern Linguistics Society (NELS)* 16, 305-321.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum.

Nishigauchi, Taisuke & Thomas Roeper. 1987. Deductive parameters and the growth of empty categories. In Roeper & Williams (eds.), 91-122.

Ouhalla, Jamal. 1990. Sentential negation, relativized minimality and the aspectual status of auxiliaries. *The Linguistic Review* 7, 183-231.

Phillips, Colin. 1995. Syntax at age two: Cross-linguistic differences. *MIT Working Papers in Linguistics (MITWPL)* 26, 325-382.

Pierce, Amy. 1989. On the emergence of syntax: A crosslinguistic study. Cambridge, MA: MIT dissertation.

Pierce, Amy. 1992. *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars* (Studies in Theoretical Psycholinguistics 14). Dordrecht: Kluwer.

Pinker, Steven. 1984. *Language Learnability and Language Development* (Cognitive Science Series 7). Cambridge, MA: Harvard University Press.

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

Rasetti, Lucienne. 2003. Optional categories in early French syntax: A developmental study of root infinitives and null arguments. Geneva: Université de Genève dissertation.

Rizzi, Luigi. 1982. *Issues in Italian Syntax* (Studies in Generative Grammar 11). Dordrecht: Foris.

Rizzi, Luigi. 1990. *Relativized Minimality* (Linguistic Inquiry Monograph 16). Cambridge, MA: MIT Press.

Rizzi, Luigi. 1993/1994. Some notes on linguistic theory and language development: The case of root infinitives. *Language Acquisition* 3(4), 371-393.

Rizzi, Luigi. 2005. On the grammatical basis of language development: A case study. In Guglielmo Cinque & Richard Kayne (eds.) *The Oxford Handbook of Comparative Syntax* (Oxford Handbooks), 70-109. New York: Oxford University Press.

Roeper, Thomas. 1999. Universal bilingualism. *Bilingualism: Language and Cognition* 2(3), 169-186.

Roeper, Thomas & Jill de Villiers. 1991. Ordered decisions in the acquisition of *wh*-questions. In Helen Goodluck, Jürgen Weissenborn & Thomas Roeper (eds.), *Theoretical Issues in Language Development*: *Continuity and Change in Development*. Hillsdale, NJ: Erlbaum.

Roeper, Thomas & Edwin Williams (eds.). 1987. *Parameter Setting*. Dordrecht: Reidel.

Saffran, Jenny. 2001. The use of predictive dependencies in language learning. *Journal of Memory and Language* 44(4), 493-515.

Saffran, Jenny. 2002. Constraints on statistical language learning. *Journal of Memory and Language* 47(1), 172-196.

Saffran, Jenny, Richard Aslin & Elissa Newport. 1996. Statistical learning by 8-month old infants. *Science* 274(5294), 1926-1928.

Schütze, Carson & Kenneth Wexler. 1996. Subject case licensing and English root infinitives. *Boston University Conference on Child Language Development (BUCLD)* 20(2), 670-681.

Snyder, William. 2001. On the nature of syntactic variation: Evidence from complex predicates and complex word-formation. *Language* 77(2), 324-342.

Soderstrom, Melanie, Kenneth Wexler & Peter Jusczyk. 2002. English-learning toddlers' sensitivity to agreement: Morphology in receptive grammar. *Boston University Conference on Child Language Development (BUCLD)* 26(2), 643-652.

Stromswold, Karin. 1990. Learnability and the acquisition of auxiliaries. Cambridge, MA: MIT dissertation.

Tesan, Graciela. 2005. What do children have in their heads? Functional heads and parameter setting in child language. College Park, MD: University of Maryland dissertation.

Tesan, Graciela & Rosalind Thornton. 2004. Overspecification in early English. In Jacqueline van Kampen & Sergio Baauw (eds.), *Proceedings of GALA 2003 (Generative Approaches to Language Acquisition Conference)*, vol. 2 (LOT Occasional Series 3). Utrecht: LOT, 469-480.

Thornton, Rosalind. 1990. Adventures in long-distance moving: The acquisition of complex wh-questions. Storrs, CT: University of Connecticut dissertation.

Thornton, Rosalind. In press. Why continuity. *Natural Language & Linguistic Theory*.

Travis, Lisa. 1984. Parameters and the effects of word order variation. Cambridge, MA: MIT dissertation.

Valian, Virginia. 1991. Syntactic subjects in the early speech of American and Italian children. *Cognition* 40(1–2), 21-82.

Vikner, Sten. 1995. *Verb Movement and Expletive Subjects in the Germanic Languages* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Wexler, Kenneth. 1994. Optional infinitives, head movement, and the economy of derivation in child language. In David Lightfoot & Norbert Hornstein (eds.), *Verb Movement*, 288-315. Cambridge: Cambridge University Press.

Wexler, Kenneth. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106(1–4), 23-79.

Wexler, Kenneth & Rita Manzini. 1987. Parameters and learnability in binding theory. In Roeper & Williams (eds.), 41-76.

Wijnen, Frank. 1997. Temporal reference and eventivity in root infinitivals. In Jeannette Schaeffer (ed.), *The Interpretation of Root Infinitives and Bare Nouns in Child Language* (MIT Occasional Papers in Linguistics 12). Cambridge, MA: MITWPL, 1-25.

Witkoś, Jacek. 1998. *The Syntax of Clitics: Steps towards a Minimalist Account*. Poznań: Motivex.

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

Yang, Charles. 2004. Universal grammar, statistics, or both? *Trends in Cognitive Science* 8(10), 451-456.

Zwicky, Arnold M. & Geoffrey K. Pullum. 1983. Cliticization vs. inflection: English *n't*. *Language* 59(3), 502-513.

Rosalind Thornton
Macquarie University
Linguistics Department / MACCS
North Ryde
Sydney, NSW 2109
Australia
rosalind.thornton@ling.mq.edu.au

Graciela Tesan
Macquarie University
Macquarie Centre for Cognitive Science
North Ryde
Sydney, NSW 2109
Australia
gtesan@maccs.mq.edu.au

Clarifying the Notion “Parameter”

Juan Uriagereka

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

Keywords: learnability; minimalism; optimality; parameter

1. Classes of Variation

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

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

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

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

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

ISSN 1450–3417
Biolinguistics 1: 099–113, 2007
http://www.biolinguistics.eu

(2) a. Nik maite nuen Lucy.
lit. 'I love Lucy.' (OK in English)

b. Lucy nik maite nuen.
lit. 'Lucy I love.' (OK in English only with topicalization)

c. Nik Lucy maite nuen.
lit. 'I Lucy love.' (Permutations OK too, all * in English)

(3) a. Who has John seen [pictures of *t*] ?

b. * Noren ikusi ditu [*t* argazkiak] Jonek?
who.GEN see III.AUX.III pictures.ABS Jon.ERG
'Of whom has Jon seen pictures?'

(4) a. [Pictures of whom] has John seen *t* ?

b. [Noren argazkiak] t ikusi ditu Jonek?
who.GEN pictures.ABS see III.AUX.III Jon-ERG
'Pictures of whom has Jon seen?'

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

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

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

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

(5) a. Juan la está viendo (a María). *Spanish*
Juan CL is seeing to María
'Juan is seeing María.'

b. John is seeing'er (*Mary). *English*

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

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

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

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

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

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

---

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

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

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

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

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

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

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

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

2. A Case Study

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

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

(cf. **Who John has seen* t ?)

b. [A quién [vio [Pedro [*t* [*t* *t*]]]]] ?
to whom *saw* *Pedro*

(cf. **A quién Pedro vio?*)

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

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

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

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

(10) a. [I wonder $[_{\text{CP}}$ who [John [has [seen *t*]]]]]
b. … $[_{\text{CP}}$ C^0 $[_{\text{IP}}$ John [has $[_{\text{VP}}$ seen who]]]]

(10b) goes back in the derivation to the point prior to the displacement of *who*. Suppose all XPs along the way are potential barriers for this displacement. VP is in construction with a lexical head, its own; as a result, VP is not a barrier for the movement of *who*. Of course, by this sort of reasoning, all categories headed by a substantive head will not be barriers to displacement. If the I–, or T(ense)–, head of IP also counts as substantive, then IP will not be a barrier either. Now CP is headed by the abstract C^0, not a substantive element. But is it in construction with a selecting element? It is, if the CP is selected by *wonder*. Then CP is not a barrier either, though not for intrinsic reasons (its own head), but by way of its contextual properties. This is the step missing in (8a), where nothing selects the relevant CP. But this CP may cease to be a barrier if it incorporates a neighboring lexical head, in which case the combined projection will be in construction with an appropriate L-marker. That's what head movement arguably achieves:

(11) $[_{\text{CP}}$ has–C^0 $[_{\text{IP}}$ John [*t* $[_{\text{VP}}$ seen who]]]]

Observe that the displacement of the verb in (11) does not cross the CP but adjoins instead to C^0. In contrast, *who* must cross CP; however this category is no longer a barrier after verb movement, in the manner indicated. The reasoning is rounded up by the assumption that the mechanism is, in some relevant sense,

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

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

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

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

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

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

(13) a. Por qué [$_{CP}$ Juan vio a María] *t* ?
why Juan saw to María
'Why Juan saw María?'

b. Por qué vio [$_{CP}$ Juan a María] *t* ?
'Why did Juan see María?'

c. * Por qué [$_{CP}$ tú pensaste que [$_{CP}$ Juan vio a María] *t*] ?
why you thought that Juan saw to María
'Why you thought that Juan saw María?'

d. Por qué pensaste [$_{CP}$ tú que [$_{CP}$ Juan vio a María] *t*] ?
'Why did you think that Juan saw María?'

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

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

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

As Chomsky asked nearly 30 years ago:

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

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

3. Considerations about Syntactic Change

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

---

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

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

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

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

---

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

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

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

To be concrete, reconsider adjunct questions involving a verbal inversion. If this innovation is peripheral, speakers should fall into it *under finite-state conditions involving adjacency*. This is plausible: The moved Wh-element is in the specifier of CP, while the putative verb movement carries the verb to the C–head, thus to a position which, in phonetic form, ends up being right-adjacent to the specifier. This is the sort of window that a surfacey finite-state analysis can see through. The consequence is interesting. Although, as we saw, Universal Grammar (UG) would not require the displacement of the verb to C^0 in relevant adjunct questions, *a costly movement is allowed in order to meet the string syntax of other structures* involving (argument) *wh*-elements and adjacent verbs. In other words, speakers are not aware of why they move a verb to the domain of adjacency of an argument *wh*-phrase; however, they can, in effect, be made aware, due to some form of more or less subtle societal pressure, that the relevant output has been achieved and that it ought to generalize. That conclusion is not demanded by UG — but it is consistent with it (disregarding a preference for absence of move-ment).

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

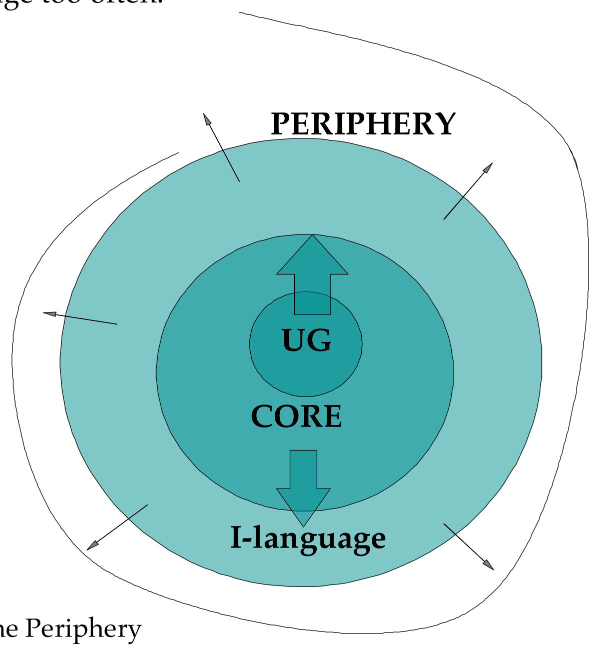


Figure 1: From I-language to the Periphery

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

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

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

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

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

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

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

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

4. Where Does This Leave Minimalism and Optimality?

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

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

---

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

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

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

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

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

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

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

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

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

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

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

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

References

Baker, Mark. 1996. *The Polysynthesis Parameter* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.

Cornell, Thomas L. & James Rogers. 1999. Model theoretic syntax. In Lisa L.-S. Cheng & Rint Sybesma (eds.), *The First Glot International State-of-the-Article Book* (Studies in Generative Grammar 48), 171-198. Berlin: Mouton de Gruyter.

Crain, Stephen & Rosalind Thornton. 1998. *Investigations in Universal Grammar: A Guide to Experiments on the Acquisition of Syntax and Semantics* (Language, Speech, and Communication). Cambridge, MA: MIT Press.

Fodor, Janet Dean. 1998. Unambiguous triggers. *Linguistic Inquiry* 29(1), 1-36.

Jarvis, Erich D. 2006. Evolution of brain structures for vocal learning in birds: A synopsis. *Acta Zoologica Sinica* 52(Supplement), 85-89.

Joshi, Aravind K. 1985. How much context-sensitivity is necessary for characterizing structural descriptions — Tree Adjoining Grammars. In David Dowty, Lauri Karttunen & Arnold M. Zwicky (eds.), *Natural Language Processing: Theoretical, Computational, and Psychological Perspectives* (Studies in Natural Language Processing), 206-250. Cambridge: Cambridge University Press.

Kroch, Anthony. 2000. Syntactic change. In Mark Baltin & Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory* (Blackwell Handbooks in Linguistics), 629-739. Malden, MA: Blackwell.

Labov, William. 1994. *Principles of Linguistic Change*, vol. II: *Social Factors* (Language in Society). Oxford: Blackwell.

Lasnik, Howard & Juan Uriagereka. Forthcoming. Structure. To appear in Martin Stokhof & Jeroen Groenendijk (eds.), *Handbook of the Philosophy of Linguistics* (Handbook of Philosophy of Sciences). Amsterdam: Elsevier.

Lightfoot, David. 1999. *The Development of Language: Acquisition, Change, and Evolution* (Blackwell-Maryland Distinguished Lecture Series on Language and Cognition 1). Malden, MA: Blackwell.

Moyna, Maria Irene. 2007. *Voseo/Tuteo* replacement in Rio de la Plata Spanish: Saliency and learnability effects on paradigms. Paper presented at the *18th International Conference on Historical Linguistics (ICHL 18)*, Montreal. [Université de Québec à Montréal, 6-11 August 2007.]

Piattelli-Palmarini, Massimo & Donata Vercelli. In press. Language in an epigenetic framework. In Massimo Piattelli-Palmarini, Juan Uriagereka & Pello Salaburu (eds.). *Of Minds and Language: The Basque Country Encounter with Noam Chomsky*. New York: Oxford University Press.

Rizzi, Luigi. 1990. *Relativized Minimality* (Linguistic Inquiry Monograph 16). Cambridge, MA: MIT Press.

Snyder, William. 2007. *Child Language: The Parametric Approach*. New York:

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

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

★ BRIEFS ★

A Note on the Default Values of Parameters

Koji Sugisaki

Modern linguistic theory attempts to explain why language acquisition is possible despite the fact that relevant experience available to children is severely limited (coined "Plato's problem" in Chomsky 1986). The proposed answer postulates that a human child is genetically equipped with Universal Grammar (UG), the initial state of the language faculty that narrowly constrains the space of hypotheses to entertain. Under the Principles–and–Parameters (P&P) approach, UG consists of (i) a number of principles that specify the properties to be satisfied by any language and (ii) a finite collection of parameters that sharply restricts the range of possible cross-linguistic variation. Chomsky (1995: 6) argues that within this framework, the task for a child in acquiring her native language is to identify the correct settings of parameters for the community's language, as stated in (1).

(1) [Within the P&P approach — KS], language acquisition is interpreted as the process of fixing the parameters of the initial state in one of the permissible ways.

Chomsky's statement in (1) can be construed as claiming that parameter-setting is the most significant factor in explaining the observable changes in the course of acquisition.

Chomsky (2004: 104) makes an additional assumption about parameters. He suggests that all parameters of UG have a *default setting*, and are specified for certain settings prior to any linguistic experience.

(2) At S_0 [initial state — KS], all parameters are set with unmarked values.

Given these two fundamental assumptions proposed by Chomsky, the following question can be raised: Is (2) compatible with (1)? I suggest that the answer is negative.

If every parameter has a default setting, then there should be a particular grammar that corresponds exactly to the initial state (UG): There should be a language (among the possibilities permitted by UG) whose parameter settings are completely the same as the default settings. Then, when a child acquires this language, no change of parametric values would take place, for the very reason that the default values are exactly the target values. Since the correct settings are

I would like to thank Cedric Boeckx, Noam Chomsky, Yukio Otsu, and William Snyder for valuable discussion. The usual disclaimers apply.

ISSN 1450–3417

Biolinguistics 1: 114–117, 2007
http://www.biolinguistics.eu

there from the start, the acquisition of lexical items (and probably, maturation of some UG properties) would be the only developmental factor that induces observable consequences in its acquisition.

Note that even in the acquisition of this language, a child has to figure out whether each of these default values is in fact the correct one for the target language. Yet, this process itself would not yield any observable change in this case: There is acquisitional evidence suggesting that the effects of default settings (as well as those of the target settings) are reflected in the child's utterances. Null subjects and '*wh*-copying' constructions in child English, illustrated in (3), are well-known phenomena that have been explained in terms of default settings of parameters.

(3) a. ____ want more apples. (Hyams 1986)
b. <u>What</u> do you think <u>what</u> Cookie Monster eats?
(Thornton 1990, McDaniel *et al.* 1995)

The default values relevant to (3) are incorrect for adult English, and hence the shift from the default to the target setting yields observable effects in the acquisition of English. In contrast, since these default specifications are correct for adult Spanish (in the case of null subjects) and for adult Romani (in the case of '*wh*-copying' constructions), the shift from the default to the target setting would not yield any observable change in the acquisition of these languages.

The above discussion leads to the conclusion that if all the parameters are specified for a default, there can be a language in which parameter setting plays no role in explaining the observable changes in the course of acquisition. If the proposal in (1) should be interpreted as the claim that parameter-setting is the most significant factor in explaining the observable changes in the course of acquisition, the assumption in (2) is not compatible with this claim, because the hypothesis in (2) permits a language in which parameter-setting induces no observable consequence in its acquisition. In order to maintain the fundamental idea in (1), we should abandon (2) and instead adopt a weaker assumption: There are parameters without any default specification, and with respect to these parameters, none of their values are employed until the child determines the correct settings for her target grammar. The evidence from the acquisition of preposition-stranding (P-stranding) and pied-piping reported in Sugisaki & Snyder 2003 in fact suggests that this weaker hypothesis is on the right track.

Cross-linguistically, the possibility of P-stranding in *wh*-questions is among the more exotic properties of English: While P-stranding is possible in English and in Scandinavian languages, pied-piping of prepositions is obligatory in most other languages (see the examples from Spanish in (5)). Given this cross-linguistic variation, a number of syntactic analyses (including Hornstein & Weinberg 1981, Kayne 1981, Law 1998, and Stowell 1981) have proposed a parameter with two values, one leading to the availability of P-stranding, and the other leading to obligatory pied-piping.

(4) What did they talk <u>about</u> *t* ?

(5) *Spanish*

a. *Cuál asunto hablaban sobre *t* ?
which subject were.they.talking about
'Which subject were they talking about?'

b. Sobre cuál asunto hablaban *t* ?
about which subject were.they.talking
'About which subject were they talking?'

If the parameter of P-stranding consists of two values, and if every parameter is specified for a default setting, then one of the following two predictions should hold with respect to the acquisition of P-stranding:

(6) a. *Prediction A*

If the P-stranding value is the default, then children learning either English *or* Spanish should use P-stranding when they first begin to apply *wh*-movement to prepositional objects.

b. *Prediction B*

If the pied-piping value is the default, then children learning English should pass through a pied-piping stage before they begin to use P-stranding.

Sugisaki & Snyder (2003) evaluated these two predictions, by analyzing the spontaneous speech data of ten English-learning children and four Spanish-learning children, selected from the CHILDES database (MacWhinney 2000). The results falsified both of these predictions. As for English, six children acquired direct-object *wh*-questions significantly earlier than P-stranding. In the utterances of these six children, no example of pied-piping appeared before the acquisition of P-stranding. As for Spanish, no single example of P-stranding was observed in children's utterances.

These results indicate that Spanish-learning children do not pass through a P-stranding stage before they acquire pied-piping, much as English-learning children do not pass through a pied-piping stage before they acquire P-stranding. Thus, children's acquisition of P-stranding in English and of pied-piping in Spanish provides clear evidence that not every parameter has a default specification: Neither pied-piping nor P-stranding is employed until the child determines the correct setting for her target grammar.

In sum, the hypothesis in (2) faces both conceptual and empirical problems. As far as I can see, there seems no reason to assume that UG corresponds to a particular grammar.

References

Chomsky, Noam. 1986. *Knowledge of Language: Its Nature, Origin, and Use* (Convergence). New York: Praeger.

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and Beyond*, vol. 3: *The Cartography of Syntactic Structures* (Oxford Studies in Comparative Syntax), 104-131. New York: Oxford University Press.

Hornstein, Norbert & Amy Weinberg. 1981. Case theory and preposition stranding. *Linguistic Inquiry* 12(1), 55-91.

Hyams, Nina. 1986. *Language Acquisition and the Theory of Parameters*. Dordrecht: Reidel.

Kayne, Richard. 1981. On certain differences between French and English. *Linguistic Inquiry* 12(3), 349-371.

Law, Paul. 1998. A unified analysis of P-stranding in Romance and Germanic. *North Eastern Linguistics Society* (*NELS)* 28, 219-234.

McDaniel, Dana, Bonnie Chiu & Thomas L. Maxfield. 1995. Parameters for *wh*-movement types: Evidence from child English. *Natural Language & Linguistic Theory* 13(4), 709-753.

MacWhinney, Brian. 2000. *The CHILDES Project: Tools for Analyzing Talk* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum.

Stowell, Timothy. 1981. Origins of phrase structure. Cambridge, MA: MIT dissertation.

Sugisaki, Koji & William Snyder. 2003. Do parameters have default values?: Evidence from the acquisition of English and Spanish. In Yukio Otsu (ed.), *The Proceedings of the Fourth Tokyo Conference on Psycholinguistics*, 215-237. Tokyo: Hituzi Syobo.

Thornton, Rosalind. 1990. Adventures in long-distance moving: The acquisition of complex wh-questions. Storrs, CT: University of Connecticut dissertation.

Koji Sugisaki
Mie University
Faculty of Humanities and Social Sciences
1577 Kurimamachiya-cho
Tsu, Mie 514-8507
Japan
sugisaki@human.mie-u.ac.jp

★ REVIEWS ★

On Evolutionary Phonology

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

by Bridget Samuels

As its name suggests, Evolutionary Phonology (EP; Blevins 2004) finds parallels in evolutionary biology. Blevins makes several mentions of Darwinian principles such as adaptation and in many cases utilizes language and reasoning that would not seem out of place in the evolutionary biology literature. However, she cautions that parallels to Darwin are necessarily "largely metaphorical" because phonological systems are learned, not transmitted in the DNA (xi). Here I think Blevins gives herself too little credit. Salikoko Mufwene (2001 *et seq*.) has made convincing arguments that languages can and should be viewed as species, with idiolects parallel to individual organisms, and that "[this] approach is analogical only to the extent that it is inspired by scholarship on biological evolution" (Mufwene 2005: 30). Certainly, the evolutionary jargon Blevins applies to linguistics is no more metaphorical than other such terms already in wide use, such as "genetic relatedness." Elsewhere, such as in chapter 2's discussion of adaptation, the correct parallels with biology are less obviously helpful, as Blevins notes. She speaks of multiple dimensions of adaptativeness — one for ease of acquisition, one for ease of articulation, and one for ease of communication — but downplays the roles of adaptation, non-aptation, and disaptation in sound change. Probing this further could provide an interesting avenue of research; one gets the feeling that the story here may not be a simple one.

The fundamental tenets of EP resonate with arguments made by the Neogrammarians, Otto Jespersen, Joseph Greenberg, and particularly Baudouin de Courtenay. These founding fathers of phonology were adamant that synchronic sound systems are best understood through the changes that produce them. Blevins also espouses this principle but differs from the tradition by rejecting teleology in sound change. For her, the only goal-directed processes that interact with pure phonological change are morphological analogy and the pressure to preserve paradigms where adhering to a regular sound change would cause paradigmatic contrasts to collapse. The elimination of teleology from phonology provides one way in which EP differs from the currently dominant paradigm in synchronic phonology, Optimality Theory (OT; Prince & Smolensky 1993).

EP and OT also clash in another closely related domain, namely how the relative frequencies of various sound patterns should be explained. In OT,

biolinguistics
© The Author (2007)
ISSN 1450–3417
Biolinguistics 1: 118-122, 2007
http://www.biolinguistics.eu

constraints on synchronic grammars and cross-linguistically fixed rankings of such constraints serve to create a markedness hierarchy. The more marked a sound pattern, the rarer it will be. In contrast, EP treats markedness as an epiphenomenon — an E-language concept belonging strictly to the domain of performance, not competence. Under this conception of phonology, because some sound changes are rare, the synchronic patterns created by those changes will also be rare. Another reason why some sound patterns are rare is that multiple independent sound changes must occur sequentially in order for those patterns to arise. Patterns formed by common changes or sets thereof will occur at a higher frequency than patterns necessitating rarer chains of events. Thus, the work done by synchronic constraints in OT instead falls upon the language acquisition mechanism, which itself drives phonological change. Understanding diachronic phonology thus requires synthesizing research from such disparate domains as phonetics, auditory perception, language acquisition, typology, dialectal/idiolectal variation, experimental phonology, and phonological theory.

Apart from simply making use of diachronic phonology to explain synchronic patterns, Blevins also proposes a new model of sound change itself. All phonetically-motivated sound change falls into one (or more) of three categories in the 'three-C' model of EP: CHANGE, CHANCE, and CHOICE. The first of these, CHANGE, covers the range of cases in which a learner mishears an utterance and treats it as a token of a different but perceptually similar utterance. An example of CHANGE that Blevins gives is the sequence /anpa/ being misinterpreted as /ampa/ due to the weakness of the cues indicating the place of the pre-consonantal nasal.

CHANCE changes are those in which the hearer reconstructs an underlying representation of an inherently ambiguous signal which differs from that of the speaker. A hypothetical instance of CHANCE would involve [ˀa̰ˀ] being analyzed as /ˀa/, /aˀ/, /ˀa̰ˀ/, or /a̰/, provided this representation differs from what the speaker has in mind. Frequency guides the analysis, so less frequent sequences are less likely to be posited as underlying forms. Language-specific constraints, which themselves must be learned in the EP model, may also come into play here, though I do not see how the phonologist (or the child) can determine when to posit a constraint and when doing so would be redundant restatement of a generalization better left as emergent. The Feature–to–Segment Mapping Principle, a property of the acquisition process with OCP-like effects, also affects CHANCE, leading the learner to assume a single source for a single phonetic feature, disadvantaging a multiple-source analysis like /ˀa̰ˀ/. The result of CHANCE is imperceptible, entailing no immediate change in pronunciation.

CHOICE, in contrast, produces tiny shifts in pronunciation akin to those documented in the Labovian tradition. When there are multiple variants of an utterance in circulation and the hearer adopts a phonological representation or "best exemplar" that differs from the speaker's, this is an instance of CHOICE. Upon hearing [kkáta] in alternation with [kăkáta] and [kakata], a listener could assume underlying /kkáta/ and an epenthesis rule, rather than the speaker's underlying /kakata/ with a vowel shortening/deletion rule. In none of these three types of sound change do we see ease of articulation or ease of pronunciation directly influencing the direction of change. Instead, like markedness, these are taken to be emergent properties.

EP's three-C model of sound change feels intuitive in some respects. It is hard to argue that something very much like CHANCE, CHANGE, and CHOICE do not play any role in sound change. However, it is less clear that they are the *only* players: Explaining how these mishearings of individual words eventually explain Neogrammarian-style exceptionless sound change would not be a trivial task. It is not enough simply to say that completed sound changes undergo lexical diffusion (p. 260). Nor is it readily apparent that distinguishing among these particular three categories elucidates anything. There seems little hope of ascertaining which C has operated to produce a specific change, either *a priori* or in practice. And if this cannot be done, then the categories are deprived of utility or individual character.

Another dichotomy emphasized in the book, the distinction between 'natural' and 'unnatural' or 'crazy' phonology, could use clarification. On several occasions Blevins switches between discussion of unnatural *rule types* and unnatural *sound patterns*, which are quite separate matters. A strange historical development can in theory give rise to a well-behaved synchronic system, just as one or more natural phonological changes in the history of a language can produce sound patterns that seem unusual. In section 3.1 we are told that "this contrast [between natural and unnatural rule types — BS] is central to Evolutionary Phonology" (p. 71), but this thread is left dangling and a few pages later we are led to the conclusion that distinguishing between natural and unnatural sound patterns "seems unwarranted and indeed misguided" (p. 78). This is indeed a coherent position insofar as rules (be they diachronic or synchronic) and sound patterns can be divorced, but the discussion could perhaps benefit from making the discussions of synchronic and diachronic naturalness explicitly distinct.

In many ways, EP represents an original research program and makes predictions that differ from those made by previous theories of synchronic and diachronic phonology. Nevertheless, in at least one respect, it runs into familiar problems faced by other models. The EP theory of chain shifting largely resembles the one put forward by William Labov (1994), for better or for worse. Both make use of exemplar theory to advance sound change, and both take the gradualness of sound change and the continuousness of the vowel space seriously. I quote below from a summary of EP position on chain shifting:

> Vocalic chain shifts are the combined result of intrinsic variation with the prototype structure of vocalic categories. Chain shifts can arise naturally when a formerly occupied area of the psycho-acoustic space is opened up, with variation giving rise to better prototypes of a pre-existing category in the newly opened space. (Blevins 2004: 291)

Such a model fails to account for any shift in which one vowel moves anywhere other than the midway point between its neighbors, as must happen in the "encroachment" commonly thought to cause push chains. These limitations fall naturally out of the particular type of exemplar theory that Blevins adopts.

The basis of Blevins' mechanism is Pierrehumbert's (2001) exemplar-based model of perception, which EP uses to account for production facts as well. In this model, when a speaker wants to produce a vowel — [u], for the sake of argument — he attempts to produce the "best" exemplar of [u] that he has heard. Crucially, "best" in this context means *most likely to be categorized as [u]*. This statistic, Pierrehumbert claims, comes from the exemplar's summed similarity to other exemplars of [u] taken as a fraction of its summed similarity to exemplars of all vowels. Now the other members of the vowel system become critical. As a test case, consider a five-vowel system of /i, e, a, o, u/. In this vowel space, [u]'s only close neighbor is [o]. There are no vowels higher or more back than [u], and [i] is distinguished from [u] by its lack of rounding in addition to its frontness. Because of the structure of the system, the best exemplars of [u] according to the rubric would in fact be closer to the edge of the vowel space than the mean exemplar of the [u] category, because these exemplars would be least confusable with [o]. Blevins translates this perceptual effect to production by stipulating that speakers try to produce the best exemplar of a given category. This feedback loop of perception and production has the effect of distributing vowels evenly throughout the perceptual space, because the system will reach equilibrium only when the best exemplar of each category is also its mean.

Since [u] is closer in the perceptual space to [o] than it is to [i], this type of mechanism can easily model how /u/ would *begin* to front. One would in fact expect this fronting to happen in every vowel system in which [i] is not the nearest neighbor to [u]. This may explain why /u/ exhibits some degree of fronting in a wide variety of languages and dialects. More problematic, and indeed impossible to explain using solely the "best exemplar" theory of sound change, is ascertaining why /u/ would front all the way to [y] rather than stopping at the point maximally distant from both [i] and [o].

The model runs into the same problem when explaining the intermediate steps of a push chain, because such shifts would require encroachment. In other words, because vowels must always move away from one another in this model, it cannot account for situations in which /oː/ raises (almost) to [uː] prior to the fronting of /uː/ to [yː], but it *can* account for situations in which /uː/ fronts and then /oː/ later raises. Blevins unequivocally states that "it is the earlier shift of *uː* > *yː* which allows [uː] to be a potentially better exemplar of /oː/ than [oː] itself" (p. 288). This seems to predict that no chains with the same steps but opposite chronological sequence would occur. Chains involving /uː/-fronting and /oː/-raising are actually quite common, appearing so frequently in the world's languages that Labov (1994) treats them as a distinct category (his "Pattern III"). Though the literature typically calls these rephonologizations "push chains," they clearly constitute drag chains if they occurred in the order for which Blevins has an explanation. Only establishing chronologies for these shifts through independent means can tell us whether they confirm Blevins' implicit prediction or not (see Samuels 2006). If the prediction is correct and /u/-fronting always precedes /o/-raising where it occurs, as long as there is some other way to spell out why /u/ fronts all the way to /y/, Labov's Pattern III shifts can be explained.

Though I have highlighted some of the shortcomings of the EP approach here, I feel it represents an important contribution to the field, one that bridges the synchronic and diachronic admirably. The book is lucidly written, well organized, and clear in its aims. EP follows in the tradition of classic work in the field, but it is different enough from previous approaches — especially OT, the current favorite — to warrant further refinement and scrutiny. Much could be gained from pursuing this line of research. If Blevins is correct that markedness should not be represented in the competence system, this would have far-reaching consequences for synchronic phonological theory. This is an empirically testable hypothesis: To the extent that markedness-based accounts of phonological phenomena can be explained otherwise, EP finds support. Similarly, the status of processes like final voicing is critical here; whether they are impossible or just vanishingly rare makes all the difference for the theory. Surely the last word on all these matters remains to be said, but EP provides a fine place from which to start the dialogue.

References

Blevins, Juliette. 2004. *Evolutionary Phonology: The Emergence of Sound Patterns*. Cambridge: Cambridge University Press.

Labov, William. 1994. *Principles of Linguistic Change*, vol. I: *Internal Factors* (Language in Society). Oxford: Blackwell.

Mufwene, Salikoko S. 2001. *The Ecology of Language Evolution* (Cambridge Approaches to Language Contact. Cambridge: Cambridge University Press.

Mufwene, Salikoko S. 2005. Language evolution: the population genetics way. In Günter Hauska (ed.), *Gene, Sprachen und ihre Evolution*, 30-52. Regensburg: Universitätsverlag Regensburg.

Pierrehumbert, Janet B. 2001. Exemplar dynamics: word frequency, lenition and contrast. In Joan Bybee & Paul H. Hopper (eds.), *Frequency and the Emergence of Linguistic Structure* (Typological Studies in Language 45), 137-157. Amsterdam: John Benjamins.

Prince, Alan S. & Paul Smolensky. 1993. Optimality Theory: Constraint interaction in generative grammar. Ms., Rutgers State University of New Jersey, New Brunswick & University of Colorado, Boulders. [Published as *Optimality Theory: Constraint Interaction in Generative Grammar*, Malden, MA: Blackwell, 2004.]

Samuels, Bridget. 2006. Nothing to lose but their chains: Rethinking vocalic chain shifting. Cambridge, MA: Harvard AB thesis.

Bridget Samuels
Harvard University
Department of Linguistics
Boylston Hall, 3rd floor
Cambridge, MA 02138
USA
bdsamuel@fas.harvard.edu

⋆ FORUM ⋆

Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax

Jon Sprouse

1. Introduction

It almost goes without saying that acceptability judgments form a continuous spectrum. While many sentences are either clearly acceptable or clearly unacceptable, a significant number of sentences fall somewhere in between in a gray area of partial acceptability. This fact has been explicitly admitted by linguists since at least Chomsky 1965. The working assumption adopted by most linguists over the past 40 years has been that these intermediate levels of acceptability are caused by properties other than grammatical knowledge. Linguists have assumed that grammatical knowledge is categorical — sentences are either grammatical or ungrammatical — and that the continuous spectrum of acceptability is caused by extra-grammatical factors (plausibility, working memory limitations, etc.). Of course, ideas such as *strength of violation* have been introduced into theories of grammatical knowledge at various times, for instance Huang's (1982) proposal that ECP violations are *stronger than* Subjacency violations or Chomsky's (1986) proposal that each barrier crossed leads to lower acceptability. However, with the notable exception of Optimality Theory (see especially Keller 2000, 2003), these analyses have been the exception rather than the rule.

The past ten years or so have seen a major shift in attitudes toward intermediate levels of acceptability. With the increasing use of formal experimental methods for measuring acceptability — a methodology that has come to be known as experimental syntax — it has become almost trivial to detect subtle differences along a continuous spectrum of acceptability. This new power afforded by experimental syntax raises the question of whether the continuity of acceptability reflects a continuity in grammatical knowledge that should be captured by the theory of grammar, or in other words, whether the working assumption of the past 40 years should be abandoned (see especially Keller 2000, Fanselow *et al.* 2004).

While the answer to this question is ultimately an empirical one that is far from being settled, this report presents two pieces of experimental evidence for a categorical distinction between grammatical and ungrammatical sentences. The first is a direct prediction of theories that assume categorical grammaticality. The psychological claim underlying theories of categorical grammaticality is that ungrammatical sentences have no licit representation, or in other words, cannot

biolinguistics

ISSN 1450–3417

Biolinguistics 1: 123-134, 2007
http://www.biolinguistics.eu

be constructed from the available mental computations. Grammatical sentences, on the other hand, have licit representations that can be constructed from the available mental computations. This predicts that extra-grammatical factors that affect the acceptability and are predicated on the existence of a representation, such as syntactic priming (Luka & Barsalou 2005), should not affect the acceptability of ungrammatical sentences. Section 2 presents results from Sprouse (2007) that confirm this prediction: Ungrammatical sentences, in particular island violations (Ross 1967), do not show a structural priming effect.

The second piece of evidence comes from the experimental syntax technique magnitude estimation. Unlike traditional tasks such as the *yes/no*–task and the Likert scale task in which subjects must categorize their responses, magnitude estimation allows subjects to respond using the theoretically infinite continuum of values available on the positive number line (see Bard *et al.* 1996). By removing the categorization aspect of the task, one might expect that responses would no longer show any categorical distinction between grammatical and ungrammatical sentences. Contrary to this prediction, section 3 presents evidence from Sprouse (2007) that subjects impose a form of categorization on magnitude estimation responses, and that this categorization appears to correspond with the grammatical / ungrammatical distinction.

2. Syntactic Priming and Categorical Grammaticality

Syntactic priming is the facilitation of a structure through previous exposure to that structure. For instance, speakers tend to repeat structures that they have either heard or spoken recently (e.g., Bock 1986, Pickering & Branigan 1998), and readers tend to show faster reading times for structures that they have recently read (e.g., Noppeney & Price 2004, Kaschak & Glenberg 2004). Of particular interest to our present purposes is that Luka & Barsalou (2005) found that exposure to structures in a reading task increase the acceptability of those structures in a subsequent rating task. This suggests that syntactic priming can be indexed by acceptability rating tasks, or to put it another way, that acceptability ratings are affected by repetition.

Syntactic priming is the effect of one sentence on a structurally identical subsequent sentence. This entails the *possibility* of constructing the structural representation in question. From the perspective of theories of categorical grammaticality, only grammatical sentences have representations that can be constructed, therefore only grammatical sentences should show syntactic priming effects. While Luka & Barsalou (2005) demonstrate a syntactic priming effect for (by hypothesis) grammatical structures, ungrammatical structures were precluded from the priming analysis because ungrammatical sentences could not be presented to the subjects during the reading phase. Presentation of ungrammatical sentences during the reading phase could have drawn attention to the dimension of acceptability, and potentially biased the subjects prior to the acceptability rating task.

The idea of syntactic priming of acceptability judgments is well known by syntacticians concerned with acceptability judgments. It has long been anecdotally reported that judgments tend to get better over time, a phenomenon that Snyder (2000) calls *syntactic satiation*. In fact, Snyder presents evidence that naïve subjects show a satiation effect with two ungrammatical structures (*wh*-islands and Complex NP Constraint islands) in a *yes/no*–acceptability task. Prima facie, this appears to be evidence for a syntactic priming effect of ungrammatical sentences, contrary to the predictions of theories of categorical grammaticality.

Sprouse (2007) re-examines Snyder's evidence for satiation in detail. In that paper, I argue that Snyder's design introduces a confound that may be responsible for his results: Subjects may adopt a strategy in which they attempt to equalize the number of *yes*– and *no*–responses, resulting in an increase in *yes*–responses over the course of the experiment. While the evidence for this confound cannot be reviewed here (the reader is referred to Sprouse 2007 for details), the re-designed experiments that rectify this confound can. As will become apparent shortly, once the *yes/no*–strategy confound is eliminated, there is no evidence for syntactic priming of ungrammatical structures in acceptability judgment tasks, as predicted by theories of categorical grammaticality.

2.1. Rationale and Design

Theories of categorical grammaticality predict that ungrammatical sentences will not be affected by syntactic priming. To test this prediction, four experiments were conducted, one experiment for each of the following island violations:

(1) *Islands tested for syntactic priming (satiation)*

Subject island:	Who do you think the email from is on the computer?
Adjunct island:	Who did you leave the party because Mary kissed?
Wh-island:	Who do you wonder whether Susan met?
CNPC island:	Who did you hear the rumor that David likes?

Over the course of the Subject and Adjunct island experiments, subjects were exposed to 14 instances of the island violation. Over the course of the *wh*- and CNPC island experiments, subjects were exposed to 10 instances of the island violation.

Two design features were incorporated in these experiments in order to eliminate the possibility of the *yes/no*–strategy. First, the task was magnitude estimation. The theoretically unlimited number of levels of acceptability made possible by magnitude estimation decreases the possibility of subjects counting the number of each type of response. Second, the number of grammatical and ungrammatical sentences were balanced in each experiment. Even if subjects attempt to track the number of responses (perhaps with respect to the modulus sentence; see section 3 below), the distribution is already balanced, so no change should result. Judgments were collected for each repetition of the island violation. A syntactic priming (satiation) effect would result in an upward trend in acceptability, which could be confirmed statistically through linear regression. All of the subjects were undergraduates at the University of Maryland with no formal training in linguistics. The sample sizes for the experiments were 20, 24, 20, and 17, respectively.

2.2. *Results*

Following the standard data analysis procedures in Bard *et al.* 1996 and Keller 2000, responses were divided by the reference sentence and log transformed prior to analysis.[1] The means for each exposure are given in the graphs below.

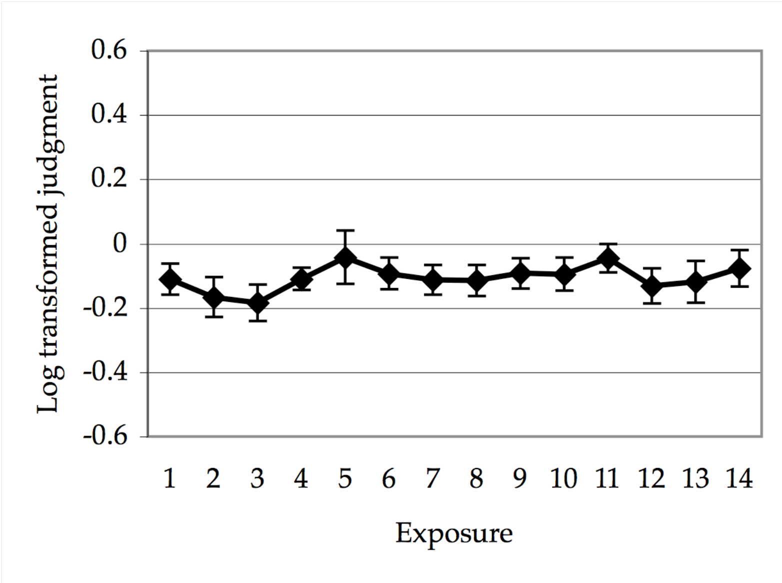


Figure 1: Subject islands

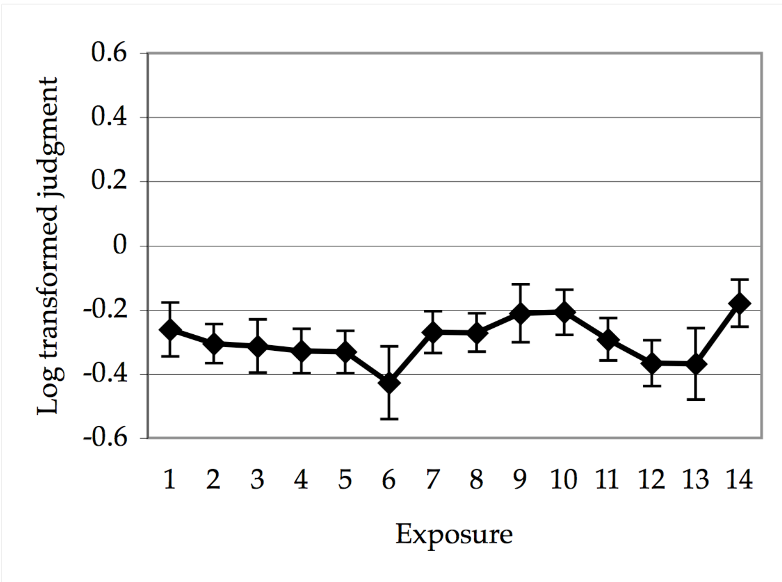


Figure 2: Adjunct islands

[1] See Sprouse 2007 for evidence that the log transformation may be unnecessary.

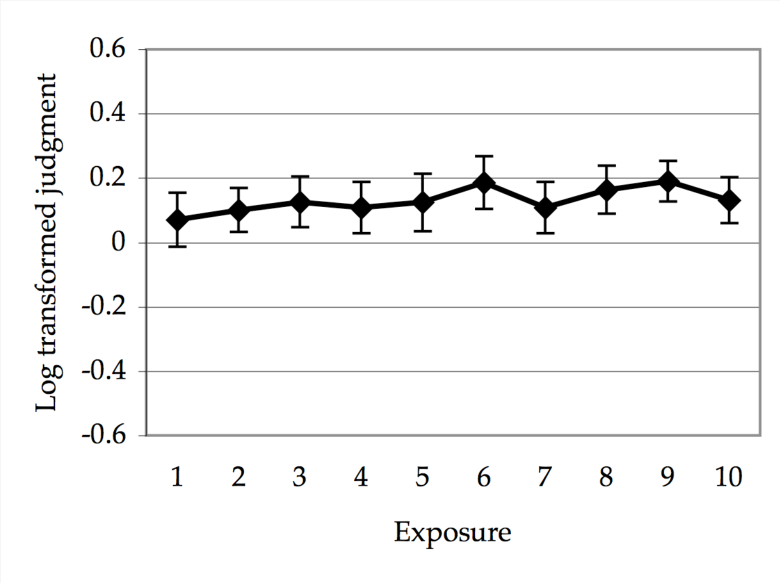


Figure 3: *Wh*-islands

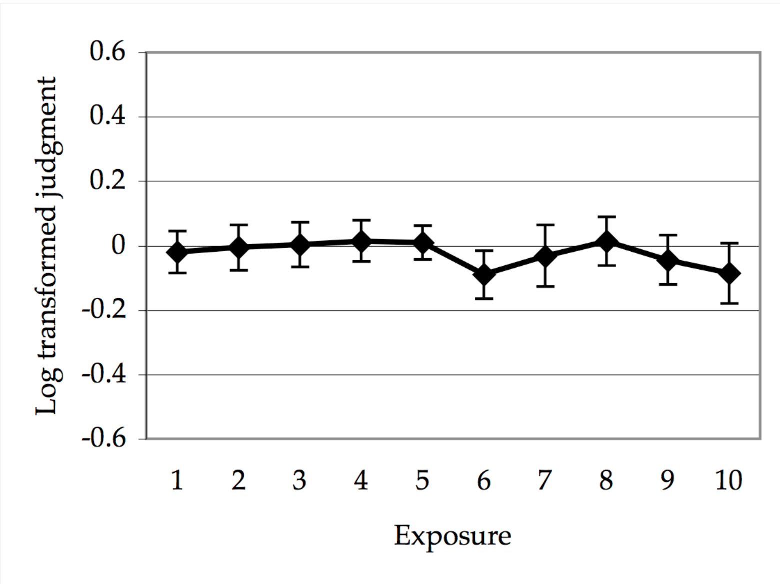


Figure 4: CNPC islands

Repeated measures linear regressions (Lorch & Meyers 1990) confirm that there are no significant increases in acceptability for any of the islands tested.

	b	*Exp(b)*	p
Subject island	–0.13	0.003	.14
Adjunct island	–0.32	0.003	.52
Wh-island	0.08	0.008	.14
CNPC island	0.001	–0.006	.44

Table 1: Linear regression of means

2.3. *Discussion*

Categorical grammaticality predicts that some extra-grammatical effects on acceptability may be asymmetrical, affecting grammatical structures but not ungrammatical structures. Syntactic priming is one such factor: The priming effect is predicated upon the existence of a licit representation. Given that ungrammatical structures have no licit representation, categorical grammaticality predicts that there should be no syntactic priming effect for ungrammatical structures. While Luka & Barsalou (2005) demonstrates syntactic priming effects on acceptability for grammatical structures, the results of these experiments suggest that there are no syntactic priming effects on acceptability for ungrammatical structures, as predicted by categorical grammaticality. These results also suggest that asymmetric extra-grammatical effects on acceptability may be a useful tool for identifying the grammatical status of structures that are neither clearly grammatical or clearly ungrammatical.

3. Magnitude Estimation and Categorical Grammaticality

In many ways, the magnitude estimation task has become synonymous with experimental syntax. The idea behind the task is simple, and perhaps easiest to explain with an example: Imagine that you are presented with a set of lines. The first line is the *modulus* or *reference*, and you are told that its length is 100 units. You can use this information to estimate the length of the other lines using your perception of visual length. For instance, if you believe that the line labeled item 1 below is twice as long as the reference line, you could assign it a length of 200 units. If you believe item 2 is only half as long as the reference line, you could assign it a length of 50 units. Figure 5 illustrates:

Reference:	———
Length:	100
Item 1:	——————
Length:	*200*
Item 2:	——
Length:	*50*
Item 3:	—————————
Length:	*300*

Figure 5: Magnitude estimation task

The resulting data are estimates of the length of the items in units equal to the length of the reference line.

Bard *et al.* (1996) were the first to argue that this technique could be adapted for the estimation of acceptability. For instance, imagine you are presented with a pair of sentences. The first is the reference sentence, and you are told that its acceptability is 100 units. The acceptability of the second sentence can then be estimated using the acceptability of the first as a reference:

Reference:	What do you wonder whether Mary bought?
Acceptability:	100
Item:	What did Lisa meet the man that bought?
Acceptability:	——

Figure 6: Estimation of acceptability

Bard *et al.* argue that this task is superior to the standard scale tasks usually used in acceptability studies because (i) it allows subjects to distinguish as many levels of acceptability as they want, unlike scale tasks in which they are not limited to 5 or 7 choices, and(ii) the distance between levels of acceptability are measured in regular units (equal to the acceptability of the reference sentence), unlike scale tasks in which the distance between points is lost to categorization.

The two benefits of magnitude estimation suggested by Bard *et al.* have in many ways become a catalyst for considering continuous models of grammatical knowledge. The freedom to distinguish a theoretically infinite number of levels of acceptability and the ability to quantify the distance between those levels with a standard unit are exactly what is needed to precisely model continuous acceptability. However, it is the underlying assumption of linearity that truly enables a continuous model of grammaticality: The magnitude estimation task makes no explicit distinction between grammatical and ungrammatical structures. They are both measured using the same reference (measured in the same units), as if they form a linear system. Sprouse (2007) investigated the linearity assumption of magnitude estimation data in detail. As will become evident below, one of the results of that study suggests that subjects introducing a categorical distinction in magnitude estimation data that appears to mirror the theoretical distinction between grammatical and ungrammatical sentences.

3.1. Rationale and Design

Two magnitude estimation experiments were conducted, each identical in content and design, except for the reference sentence. The reference sentence for the first experiment was a type of *wh*-island violation, specifically an *if*-island: *What do you wonder if Larry bought?*. The reference for the second experiment was a Coordinate Structure Constraint violation: *What do you think Larry bought a shirt and?*. The critical manipulation is that the reference sentence for experiment 2, the Coordinate Structure Constraint (CSC) violation, is one of the conditions in the experiment. Therefore the values obtained from the first experiment can be used to predict the values in the second experiment. We know that the value of the CSC in the second experiment is going to be 1 (because it is the reference unit). So for instance, if the CSC is rated as .5 in the first experiment, we would expect all of the values in the second experiment to be doubled (because .5 x 2 is 1). If the CSC is rated as .25 in the first experiment, then we would expect all of the values in the second experiment to be quadrupled (because .25 x 4 is 1). In other words, the distribution of the results will remain constant, but the absolute values of the results will be transformed by the value necessary to translate the CSC into a single unit.

The body of the experiments were identical. Each contained 8 different violations types:

(2) *Violations judged in magnitude estimation experiments*

a. *Adjunct Island*
What does Jeff do the housework because Cindy injured?

b. *Coordinate Structure Constraint*
What did Sarah claim she wrote the article and?

c. *Infinitival Sentential Subject Island*
What will to admit in public be easier someday?

d. *Left Branch Condition*
How much did Mary saw that you earned money?

e. *Relative Clause Island*
What did Sarah meet the mechanic who fixed quickly?

f. *Sentential Subject Island*
What does that you bought anger the other students?

g. *Complex NP Constraint*
What did you doubt the claim that Jesse invented?

h. Whether-*Island*
What do you wonder whether Sharon spilled by accident?

Subjects judged five of each violation type. All subjects were University of Maryland undergraduates with no formal training in linguistics. 22 subjects participated in the first experiment (*If*-reference), 31 participated in the second (CSC-reference).

3.2. Results

Following the standard data analysis procedures in Bard *et al.* 1996 and Keller 2000, responses were divided by the reference sentence and log transformed prior to analysis. The mean judgments for each violation type for both experiments are provided in the following graph:

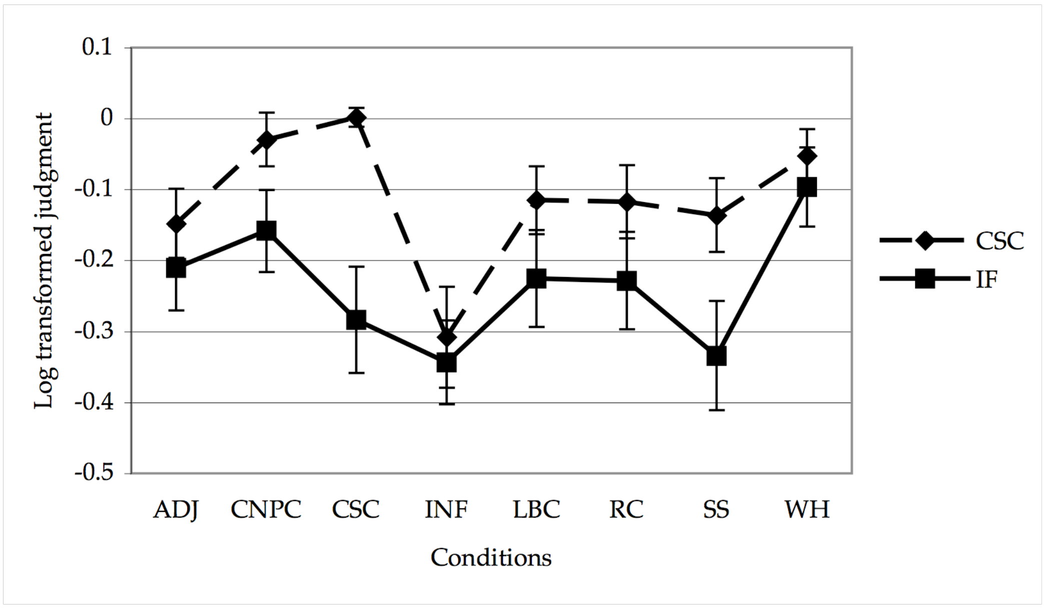


Figure 7: Mean ratings for *If* and CSC references

The first thing to notice is that the shape of the two lines is nearly identical. This suggests that the relative judgments remain constant across the two experiments. This can be quickly confirmed by comparing pairs of conditions in the two experiments. As the table below indicates, three pairs of conditions that are statistically different in the first experiment are also statistically different in the second experiment:

	If-reference		*CSC-reference*	
Contrast	t	*p*	t	*p*
ADJ-INF	3.86	.001	4.50	.001
CNPC-SS	3.58	.001	3.10	.002
Wh-RC	4.88	.001	1.70	.05

Table 2: t-Tests for *if* and CSC references

Therefore we can be confident that the two experiments have a similar level of sensitivity to acceptability.

Despite the high degree of agreement with respect to relative judgments, there is an anomaly with respect to the overall distribution of the judgments. Notice that the CSC violation is approximately in the middle of the distribution of judgments in the first (*if*-reference) experiment: The CNPC and *wh*-conditions are more acceptable than the CSC, the INF condition is less acceptable, and the rest are statistically equal. We would expect a similar distribution of acceptability in experiment 2, with the only difference being that the CSC should be equal to 0. However, what we find is that there are no conditions judged more acceptable than the CSC (none are above 0). Furthermore, only two are statistically equal to the CSC, the CNPC and *wh*-conditions, which were the two that were *more acceptable* than the CSC in experiment 1. The rest of the conditions are all *less acceptable* than the CSC. This is expected for the INF condition, but the rest of the conditions were statistically equal to the CSC in experiment 1

These results suggest that subjects are not actually performing the magnitude estimation task, but rather performing a relative rating task in which the reference sentence serves as an upper bound for ungrammatical items. This is also true of the results from experiment 1. In fact, in both experiments, the only conditions that are rated above the reference item are the grammatical fillers. This suggests that subjects are imposing their own categorization on an otherwise continuous task.

3.3. Discussion

By design, the magnitude estimation task provides subjects with a theoretically continuous (and infinite) response scale. As such, it is no more surprising that magnitude estimation tasks yield continuous measures of acceptability than it is that *yes/no*–tasks yield categorical measures of acceptability. Yet despite the continuous nature of the magnitude estimation response scale, and despite the absence of any mention of a categorical grammaticality distinction in the magnitude estimation instructions, the subjects in these experiments appear to be imposing a categorical distinction on their acceptability judgments. While the spontaneous imposition of a categorical distinction on a continuous rating scale is surprising, it is entirely consistent with a categorical approach to grammaticality.

4. Conclusion

The tools of experimental syntax have made it possible to quantify continuous acceptability with relative ease. The question this raises is whether the tools of experimental syntax have also made it possible to investigate the predictions of theories of categorical grammaticality.

This report has presented two case studies that provide support for categorical grammaticality. The first study investigated one possible prediction of categorical grammaticality: The acceptability of grammatical and ungrammatical sentences should be affected by different factors. As predicted, the study found no syntactic priming effect for ungrammatical structures, despite evidence in the literature for syntactic priming of grammatical sentences (e.g., Luka & Barsalou 2005). The second study investigated the nature of acceptability data itself, presenting evidence that subjects will impose categorical distinctions on continuous rating scales.

While these studies can only begin to address the question of the nature of grammatical knowledge, they raise the possibility of finding more evidence for categorical grammaticality in the data being collected by experimental syntax. These results argue for a closer investigation of the tasks of experimental syntax, as well as a reconsideration of the psychological consequences of categorical grammaticality in studies of the factors affecting acceptability.

References

Bard, Ellen G., Dan Robertson & Antonella Sorace. 1996. Magnitude estimation of linguistic acceptability. *Language* 72(1), 32-68.

Bock, J. Kathryn. 1986. Syntactic persistence in language production. *Cognitive Psychology* 18(3), 355-387.

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monograph 13). Cambridge, MA: MIT Press.

Fanselow, Gisbert, Caroline Féry, Matthias Schlesewsky & Ralf Vogel. 2004. *Gradience in Grammar: Generative Perspectives*. Oxford: Oxford University Press.

Huang, C.-T. James. 1982. Logical relations in Chinese and the theory of grammar. Cambridge, MA: MIT dissertation.

Kaschak, Michael P. & Arthur M. Glenberg. 2004. The construction needs learned. *Journal of Experimental Psychology: General* 133(3), 450-467.

Keller, Frank. 2000. Gradience in grammar: Experimental and computational aspects of degrees of grammaticality. Edinburgh: University of Edinburgh dissertation.

Keller, Frank. 2003. A psychophysical law for linguistic judgments. *Proceedings of the 25th Annual Meeting of the Cognitive Science Society (CogSci–03)*, 652-657.

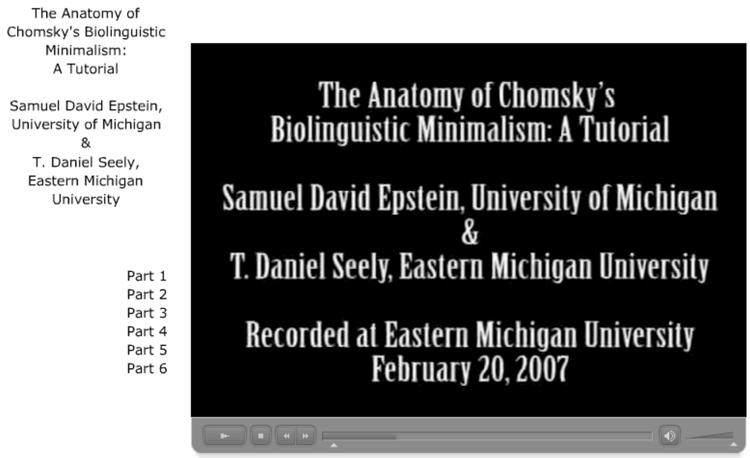
Lorch, Robert F. & James L. Meyers. 1990. Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology:*

Learning Memory and Cognition 16(1), 149-157.
Luka, Barbara J. & Lawrence W. Barsalou. 2005. Structural facilitation: Mere exposure effects for grammatical acceptability as evidence for syntactic priming in comprehension. *Journal of Memory and Language* 52(3), 436-459.
Noppeney, Uta. & C. J. Price. 2004. An fMRI study of syntactic adaptation. *Journal of Cognitive Neuroscience* 16(4), 702-713.
Pickering, Martin J. & Holly P. Branigan. 1998. The representation of verbs: Evidence from syntactic priming in language production. *Journal of Memory and Language* 39(4), 633-651.
Ross, John R. 1967. Constraints on variables in syntax. Cambridge, MA: MIT dissertation. [Published revised as *Infinite Syntax!* [Language and Being], Norwood, NJ: Ablex, 1986.]
Sprouse, Jon. 2007. A program for experimental syntax. College Park, MD: University of Maryland dissertation.
Snyder, William. 2000. An experimental investigation of syntactic satiation effects. *Linguistic Inquiry* 31(3), 575-582.

Jon Sprouse
University of California, Irvine
Department of Cognitive Sciences
3151 Social Science Plaza A
Irvine, CA 92606
USA
jsprouse@uci.edu

★ FORUM ★

The Anatomy of Biolinguistic Minimalism



Samuel David Epstein & T. Daniel Seely

URL: http://elearning.emich.edu/media/Producer/LING/SeelyEpstein.html

Most of this talk is an exposition (hopefully accurate) of Chomsky's own ideas, with a few of our own hypotheses thrown in. Our goal is to attempt to clarify some of the leading ideas of Chomsky's biolinguistic minimalism, especially those which, even as full time practitioners we have found challenging (at least for us!). We hope that the video clarifies certain issues and what we believe may be misunderstandings of Chomsky's framework and goals, including our own misunderstandings, which the making of this video helped us to clarify. We hope that our attempted clarification here will help contribute to continued biolinguistic inquiry and to continued interdisciplinary work that this framework has already generated. Certainly, none of the issues here are self-evident or simple. We have struggled to understand certain concepts, and hope we do Chomsky's theory justice here, in our attempt to convey the excitement that Chomsky's generative linguistic inquiry and biolinguistic minimalism has engendered. We have tried to be as non-technical as possible, focusing on certain of the foundational issues and current developments of the biolinguistic framework of inquiry.

This talk was presented at the Michigan Linguistics Society Annual Meeting in 2006. We thank audiences there, and at a presentation at Eastern Michigan University, for helpful comments. We are grateful to Pam Beddor, Mark Hale, and Rick Lewis for discussion, and especially to Noam Chomsky for helping to clarify various of these issues with us over the years. Thank you as well to Andy Burghardt for his tireless work on the video. Finally, we thank Kleanthes Grohmann and Cedric Boeckx for their support of this work.

biolinguistics

ISSN 1450–3417
Biolinguistics 1: 135–136, 2007
http://www.biolinguistics.eu

Key questions we'll address include:

- What is the object of inquiry in biolinguistic minimalism; and what role do idealization and abstraction play in the formation of this (and other) theories?
- What are the particular idealizations proposed?
- What methods could be used in exploring this object?
- In what sense are the methods "scientific" and in what sense is the inquiry "biological" — more specifically, can anatomy and physiology be concepts pertinent to cognitive science in general, and to linguistics construed as a cognitive science in particular?
- If so what explanatory benefits might this provide?

Biolinguistics construes the human language faculty as a 'mental organ' and so we discuss the role of "organology" in biology and biolinguistics. We then discuss methodological considerations. Here we seek to distinguish Organs vs. Behaviors vs. Capacities. The roles of Anatomy and Physiology, as well as the difference between perception and cognition, are also examined. The so-called Mind–Body problem is, again following Chomsky, argued not to be a problem unless one requires that all theoretical postulates be material tangible objects (which would exclude all of science). The theory-driven creativity of experimental laboratory science versus "indiscriminate data collection" is discussed along with the distinction between data on the one hand and evidence (for or against a specific theory) on the other. The nature of the various kinds of evidence currently used in Linguistics, including introspection as well as psycholinguistic and neurolinguistic laboratory methods, is then discussed.

Finally we turn to Minimalism, having hoped to have clarified the term "biolinguistic." We discuss two kinds of Minimalism, methodological and substantive. The first is a pervasive consideration in scientific explanation under which simplicity of explanation, like empirical coverage, is regarded as vitally important. We hope to illustrate substantive Minimalism with a very simple (we hope tractable) Minimalist syntactic analysis — which itself revisits the roles of organology, anatomy, and physiology within biolinguistic minimalism.

Samuel David Epstein
University of Michigan
Department of Linguistics
440 Lorch Hall
611 Tappan Street
Ann Arbor, MI 48109
USA
sepstein@umich.edu

T. Daniel Seely
Eastern Michigan University
Department of English Language & Literature
Program in Linguistics
612D Pray-Harrold
Ypsilanti, MI 48197
USA
tseely@emich.edu

★ FORUM ★

An Interview with Henk van Riemsdijk

by Kleanthes K. Grohmann

Platres & Zygi, Cyprus (14 October 2006)

Henk, it's mid October 2006. Let's pretend it's already 2007 — the year this interview will be printed in the new online journal Biolinguistics*. Before we talk about you, your visit to Cyprus, your work in linguistics, what comes to your mind when you hear these keywords:* 2007, new online journal, *and* biolinguistics*?*

One thing that, in retrospect, strikes me about *Syntactic Structures* [Chomsky 1957], the book that got me into generative grammar, is that it may well be the one publication by Chomsky that is least concerned, at least overtly, with mentalist issues. It seems only fitting that 2007, the 50th anniversary of *Syntactic Structures*, coincides with the first appearance of the new online journal *Biolinguistics*.

That's right. You mentioned this in the autobiographical introduction to your talk on Wednesday night [van Riemsdijk 2006c], when you said that you looked for some historical developments of generative grammar, in particular the possible genesis in Chomsky's writing of what we would now call the "biolinguistic" approach to language. Could you please briefly rehash this for a wider audience?

The following is a heavily shortened, edited, and amended version of the original interview. It basically covers the third part of the interview of which a fuller version can be obtained from http://www.punksinscience.org/kleanthes/BL/HvR.htm — but that is still an edited version (expletives and personal issues deleted); for legal reasons, the spoken corpus can *not* be made available to a wider public. The amendments to this interview are obvious: In the interest of the reader who may not be a specialist on all issues addressed, I added references to the literature as we mention them.

ISSN 1450–3417

Biolinguistics 1: 137–149, 2007
http://www.biolinguistics.eu

Well, what I said in the talk was that only two years after *Syntactic Structures*, he came up with the Skinner review [Chomsky 1959] and that was, as far as I am aware of, the first clear sign that Chomsky was adopting that frame of mind. I sort of said jokingly that in retrospect, if you tell Chomsky that he only came up with that stuff later, he would probably deny it and say that it had been clear to him right from the start — and, you know, who am I to say that he would be lying? All I'm saying is there was no real evidence in the writing that that was the main goal he was pursuing. It would actually be interesting at some point to ask him this question.

And you also mentioned that you looked for some specific keywords, like 'cognitive', 'biological', and you found them in LSLT *[Chomsky 1955] — but only in the introduction to the published edition, which Chomsky actually wrote in 1975.*

Yeah, that's right. [But bear in mind that] the publication of *Syntactic Structures*, which was like a brief summary of *LSLT*, was in part specifically designed to get his career at MIT going, and as we well know from that time, much of the research was financed through the military, and what the military were interested in was information theory. And so, clearly what he emphasized was the stuff about the Chomsky hierarchy [Chomsky 1956], things like that. I mean, if he were to say, "Well, I didn't want to endanger the success of linguistics at MIT by throwing in this psychologizing stuff that the real hard scientists are not necessarily gonna be very enchanted by" — that in psychology, in the dominant views on psychology at the time, were absolutely abhorrent — who am I to say again that that wasn't the best strategy?

In the last few years the term 'biolinguistics', has become one of those buzzwords, if you like — of course, the term has been around for long, much much longer [for a critical query in this direction, see, for example, http://linguistlist.org/issues/18/18-1379.html#2; see also the editorial to this volume, Boeckx & Grohmann 2007] — but what's your take on, let's say, the current use of the term, and possibly the work connected to it?

I think the first thing I want to say is that I'm glad that people talk about it — it shows a certain level of awareness, that this is what we're really doing all this for. The second thing is that it's actually kind of interesting that, just as we're talking more about biolinguistics, there is the "third factor" [Chomsky 2005] being emphasized in language acquisition, namely general principles of design. Now, general principles of design could be truly biological principles of design. But in actual fact, when you look at the examples that are discussed, they're really principles of design concerning the whole physical world. And what are the principles of design that people talk about? Economy and locality — and indeed these do seem to determine many properties of the physical world. So, one somewhat diabolical interpretation would be to say that, just as we are beginning to talk about biolinguistics, the main focus of attention shifts to non-biological principles. *[LAUGHTER.]*

This said, I think that would be too diabolical and in fact the wrong way of looking at it because what we're looking at is, of course, how laws of nature manifest themselves in biological systems. There already we face a reduction problem because we don't really know whether to talk about these things in terms of, let's say, a physical theory, with physical primitives and a physical alphabet, or whether to talk about these things in terms of biological objects and biological principles, biological alphabets, and so on; axioms, if you wish. And then, when we try to interpret biological principles of design in the role that they play in grammar, we face yet another step of reduction of this sort because we don't really know how to talk about biological systems that are dealing with DNA, with biochemistry, with cell structure, and so on, and how to connect that sort of thing with the sort of notions that we work with in linguistics.

Such as economy or locality. The displacement property of human language.

Exactly. It's quite clear that a great deal of thinking is gonna be needed before we can do anything really significant. And I'm actually hoping that whatever is going on around the world in terms of activities on the biolinguistic front, including the new journal, is going to help set the stage for such work.

That's certainly also our hope, as editors, and one of the reasons why we have people like you on the Advisory Board to discuss some strategies in the future. Right now, for example, this year [2006], a new e-journal was started in the biological world, the journal Biological Theory *[http://www./mitpressjournals.org/toc/biot] (even if backed by an established publisher), with which we hope to have some interaction for drawing papers, submissions, discussion, and so on. This ties in with what you just said. We don't really know yet how to talk about some of these things. Also, you and Riny [Huybregts] organized this conference at Tilburg University some years ago, on biolinguistic matters with all kinds of linguists, pychologists, biologists, neuroscientists, right?*

Oh yes, with lots of people. The conference was called *The Genetics of Language* [http://linguistlist.org/issues/12/12-682.html#1], and what we actually tried to do was mainly to create a bridge between several, rather independent brands of, call it, biolinguistics. For one thing, the people who work on population genetics try to find some significant answers to questions like, "Is the Black Eve Hypothesis correct or not?" — that is, that language originates in one single spot, and migrated and diverged from there on — or whether we should adhere to a multiple-origin story. The other thing was biolinguistics in the stricter sense: "Is there anything we can say about the structure of the human brain in relation to the property of grammar?" And there was particular emphasis on SLI, specific language impairment. We had some of the major players of that subfield at the conference. If I'm honest, I would say that interaction between these different groups left something to be desired. But then, I think, it is easy to be over-optimistic in these matters, and I'm still hoping that the very fact that these people interacted and participated in the same conference is in itself already a significant step in the right direction. And in fact, Lyle Jenkins produced a book that was largely the result of that conference [Jenkins 2004].

That is correct. That a lot remains to be desired in these interactions is natural, and it happens time and time again. The Genetics of Language *conference is one example and, from what I hear, something similar happened this year at the summer school in San Sebastian, where you had a bunch of geneticists, biologists, psychologists, Chomsky [Piattelli-Palmarini* et al.*, in press; see also Chomsky 2007, in this volume] — that they were talking to, rather than with, one another. I wasn't there, so I don't really know.*

Specifically, if someone came to you, a colleague, a junior colleague, maybe a senior col-league, and says, "Hey Henk, I saw your name on the advisory board of this new journal, Biolinguistics *— what do you expect to be doing there, what do you hope for?"*

The first part of the answer would simply be that I'm on the advisory board of a number of journals, and in general, I don't do anything. I get asked to do reviews of manuscripts a lot, but there is very little correlation between the journals for which I'm asked to do reviewing and the journals that I'm on the board of. Now, at the same time, I don't really like to lend my name to just anything, and the fact that my name is there minimally should be interpreted as meaning that I think it's a good idea that such a journal exists. And the reason that I think it is such a good idea is that it's a field that deserves attention, in which a lot of work should be done, and that has a chance of going somewhere. Now, I want to correct this to a certain extent, which is, that I seem to be suggesting that it's a separate field. But of course, what this means is that it just emphasizes that what we do as linguists is work that is ultimately to be taken to be part of the biolinguistic enterprise.

Excellent. That leads me straightaway to the next question. How — in the context of using the term 'biolinguistics' or anything like that — is biolinguistics today different from biolinguistics back then? Like in the '70s, when, for example, Lyle Jenkins started, or tried to start — I don't know whether you know that —a journal which was to be called Biolinguistics*, and he got support from a number of Nobel-prize winners in the natural, biological sciences [see Jenkins 2000 and this volume's editorial, Boeckx & Grohmann 2007]. You had the Harvard and Royaumont conferences in 1974, 1975, organized by Massimo [Piattelli-Palmarini 1974, 1980]. So that era and the notion of biolinguistics versus today — what is different?*

Well, first of all, I would say that in the realm of results, results that can be called 'biolinguistic' in a specific sense, there isn't that much to report. I would say that developments in SLI are at least interesting [e.g., Leonard 2004, Wexler 2004], and there are some clear indications now that there are genes that are in some way, that we largely don't understand, involved in matters that pertain to language in the brain.

A second remark that I think should be made is that it's easy to mistake the boom in studies in neuro-imaging as constituting a very significant step ahead. I'm not an expert, and I by no means read everything, but from what I have seen, there really isn't that much by way of significant progress to report. It appears that the level of resolution at which we currently are — but even if it goes much further — is still a long step away from telling us anything about the sort of interesting properties that generative grammar has found as being extremely unique to human language. Take recursion as the first thing that comes to mind — these two subfields just don't touch each other, not by a long stretch, and in fact I think we shouldn't necessarily be too optimistic that this is just a matter of time. It may just be that imaging of this type is simply not gonna tell us anything about things like binding, or subjacency, or recursion, or Merge, or whatever.

Well, there are some current attempts — I'm not sure how you see that, given that I come from Maryland, intellectually speaking: Dynamic systems and the biological world, Juan

Uriagereka's attempts of tying in the Fibonacci sequence in an interesting way as a property of language [Uriagereka 1998; see also David Medeiros' work cited below and Soschen 2007].

Yes, I recognize that there are a bunch of things floating around, and in all modesty I would include my own *XX-work among the same [read: "star double-X"; van Riemsdijk 2006c, in press]

And grafts [van Riemsdijk 2006d; see van Riemsdijk 2001, 2006a, 2006b]…

And grafts, well —

Maybe not.

No, I mean the fact that I am using a botanical metaphor for that really has nothing to do with the biolinguistic perspective. Not any more than the fact that we have been calling syntactic structures "trees" for many decades now, so I would leave that out.

*Let's talk about *XX then.*

First of all, it is something I perceive as recurring, popping up all the time, and it's something that is absolutely not unfamiliar when we look at the physical world. Now, as we know, perceptions may be very deceptive, so it's easy to be deceived, but for the time being I think there might be something significant going on, and I would consider Piattelli-Palmarini & Uriagereka's [2004] work on equating feature checking with anti-immune reactions as being in the same spirit.

Again, from the Genetics of Language *conference.*

Exactly. And yes, taken all together, there may be half a dozen proposals floating around that go beyond mere hand-waving. But frankly, they don't really go that far beyond hand-waving, if we're honest — and I would certainly include my own work on *XX [van Riemsdijk, in press]. But that doesn't mean that it shouldn't be done, because these are the first attempts to try and sort things out in certain ways, and we should just think about these things and keep communicating with each other about these things as a way to get more out of it. I have to confess that, for example, throwing out these sayings about recurring sequences in DNA [van Riemsdijk 2006c] — well, I feel quite embarrassed because I've picked this up from one or two articles that I found in the literature.

Basically, only junk DNA can have identical sequences…

… while recurring sequences occurring in significant DNA always need to have buffers — that's how I read the articles that I have read. But at the same time I have to say, those are articles that I understand only 15% of. Those are difficult, specialized fields, and I don't have a glimmer of a hope to really understand what's going on. Also, I don't really know for sure that those researchers that write this up are taken seriously in that field, so it is quite clear that the only way in which we might possibly get a little bit further, is by starting to talk with these people, but, of course, that, in turn, would imply that we manage to explain to them the sorts of phenomena in linguistics that we think might be interpreted along these lines. And, you know, then the big chasm shows up, which is similar in a sense to this chasm about the resolution of imaging techniques and what we really want to be looking for. We're not entirely sure that this chasm can ever be bridged, but that's what science is for: I mean, you can only try and hope, and maybe something will happen.

Well, I would say that Cedric [Boeckx] and I have a glimmer of a hope that the journal Biolinguistics *may contribute a little bit towards that by — hopefully, eventually, once we have some kind of reputation — approaching geneticists, for example, and fill them in on what we are trying to do and ask whether this makes any sense to them. Or what you just said, "Ok, look, I read this one paper, I understand 10, 15%, this is roughly what I get out of it. Would you agree?" — and then they might tell you, "Well, that's exactly what it is" or that it's not what it really is instead, or that these people shouldn't be taken seriously, and so on.*

Exactly, yes.

But you start this kind of dialogue, and hopefully by targeting a wider group of researchers in different fields, this might become common practice, if at all possible.

One would definitely hope so.

Now, you could, of course, argue that the relevance of grafting for the biolinguistic approach [van Riemsdijk 2001, 2006a, 2006b] — but then again you could say maybe that it's just a core syntactic approach — is to take an operation, Merge, seriously, in all its consequences.

Yes.

Yes, but then I guess it's just biolinguistic, biological, as much as Merge is.

Oh yes, absolutely, absolutely, yes. That is, of course, true.

I recall a paper by this guy from Arizona, David Medeiros, who's taken the Uriagereka-line on the Fibonacci sequence seriously and applied it to X-bar structure, and basically derives a system very similar to the X-bar system [Medeiros 2006; see also Carnie & Medeiros 2005 and Carnie et al. *2005]. Something that I tried to do in my dissertation, my book, from a different perspective, by getting away from this bare phrase structure in the sense that we get linear structure [Grohmann 2000, 2003, also Grohmann 2001 and updated versions, but see Atkinson 2007 for a critique] — something you argued for also in Wednesday's talk, that we really have to make available these positions: specifier, head complement [van Riemsdijk 2006c, in press]. These are the serious relations that we have in our phrase structure and nothing else, and you see this over and over again. We should be able to say something about it, derive it. However, it's not so clear to me that bare phrase structure necessarily does that [Chomsky 1994* et seq.*]. But then again, of course, bare phrase structure does a lot of other things very nicely, such as taking Merge to successively build our structure.*

Absolutely, yes.

Let me move on a little. You could say that GB was the explanandum of the P&P theory of English [cf. Chomsky & Lasnik 1993]. Most minimalists, Chomsky included, at least up to a certain degree, would place minimalism in the wider research agenda — that what we're trying to do is to come up with a theory, or with a set of rules, that describe the *principles and parameters of natural language. I'm not so sure how much our knowledge of principles of UG and parameters in the grammar have developed. Do you still see today's research program influenced by, or should it even be part of, the general P&P approach?*

You know, that's a pretty big question.

Well, it was supposed to be the last question.

Yes. Let me perhaps begin by saying that there are a number of aspects of this rather diverse thing called "minimalism" that I definitely think are interesting and worthwhile. That's certainly true of Merge. You know, being very selfish here, I think Merge is a great step forward.

Merge follows from grafts. [LAUGHTER.]

Second, the idea that all conditions can be reduced to interface properties, is certainly worth pursuing.

The strong minimalist thesis [Chomsky 1995, 2000, 2001, 2004].

It's the strong thesis. I am not sure how optimistic I am about success, but again, let's say, it's worth pursuing. Now, there are other respects in which I tend to feel that going from P&P to minimalism has possibly been a step back. I think that the idea that Move can be successfully and explanatorily regulated by features and feature checking of whatever comes after that, has largely failed. It was an interesting idea, essentially taking Vergnaud's idea from that infamous letter in the late '70s seriously [Vergnaud 1977] and trying to extend it to all of Move α. But the thing was — Actually, let me go even one step further. It was thought to be an almost incontrovertible case showing that it was possible to motivate certain types of Move α and to do it by means of features. Now, even there I think there is serious cause for doubt, and, you know, the main reason that I think that that might have been wrong to start with is that it ignores the notion of default case. In fact that's what we see when we look around the languages of the world, when a noun phrase is in a position where it cannot get Case —

Hanging topic, left dislocation, fragments.

Yes, exactly. Why, in English, the way of making up for lack of case is to insert *of*, — that's why a noun with an object gets *of* (as in *the destruction of the city),* so if direct object case is absorbed in passives, why can't you say *It is killed of John*? But none of that obtains and to my knowledge, nobody has ever given a satisfactory answer to that. So, this is just by way of an illustration of the fact that even the best-case scenario for triggered movements is on very shaky grounds and 90% of it is pretty much the worst-case scenario.

That was the motivation for your Triggers *conference a couple years ago [http://www.linguistlist.org/issues/13/13-705.html#1; see also Breitbarth & van Riemsdijk 2004]?*

Yes, yes. We were very egalitarian-minded, so we invited people from all camps and even more people submitted abstracts and so on — but yes, the outcome was definitely very mixed, very mixed. Of course, you can always say it's too early to evaluate and things like that, but I am tempted to say that that was a wrong move. Well, I'm not necessarily saying that going back to the old system would be the right move, but I sort of felt that, assuming that Move was really Move α, and to work on a system of constraints that would make sure that overgeneration was taken care of, was a perfectly worthwhile way of thinking about things. And there's nothing in principle that makes it incompatible with minimalism, it seems to me, because after all, we may be able to interpret or re-interpret the constraints that are needed to get Move α to work well in terms of interface properties.

Right. Now, one way of constraining Move α in GB was the modular structure — modular on two levels: First you had various GB modules (Case Theory, PRO Theorem, Theta Theory) that interacted in particular ways; second, you had written into the architecture several levels of control (call them levels of representation). So one core minimalist line of attack would be to ask: "Do we really need all this apparatus, can we reduce it?" [Hornstein 2001; see also Hornstein et al. *2005]. Now, to the extent that what you just said at the end, that you don't really see why GB (or why Move α from GB) could not really be reconciled with minimalism, you could say, "Well, in that technical implementation it can't be, because we don't want these things" (the modular structure and superfluous levels of representation). Or you could go the other way around and say, "No, we don't want them if they are really superfluous, but maybe we really need these four levels" — as Juan Uriagereka argues currently, that we do need some kind of D-structure component [Uriagereka 1999]. So, maybe these levels of representation need not be all interface-driven — or you could say that D-structure is actually an interface level, namely between the lexicon and the derivation. Or you might want to say, maybe you don't like the modular structure of GB, but there is some beauty about that, too — that it's nice to have these different levels, they hang together, they connect.*

Absolutely.

So it's not so clear that we should have to throw them out altogether. The other thing is when you said Move α was nice because it was unconstrained — you had to find ways to constrain it and Move, triggered movement for feature-checking purposes, does not seem to do the trick. However, in the very latest version of Phase Theory [Chomsky, in press], of the phase-based approach to syntax, that's pretty much what Chomsky says. He calls it now EPP. Now, the EPP does not have any intrinsic meaning, it doesn't have any meaning related to the '82 EPP [Chomsky 1982], that subject positions must be filled. It just means that we have a p-feature somewhere, an edge feature for various projections — but really he dissociates movement from checking again.

That's true.

We didn't have feature checking mechanisms in GB, but if we had them, given that everything should be done by means of government and c-command and m-command, it's perfectly conceivable to think of something like Agree to work in a GB framework, but then you just have to find some other explanation for movement. For Chomsky [2005], it's the optimal solution.

Yes. I think what you and I both seem to be saying is that one shouldn't exaggerate the discontinuity between GB or P&P on the one hand [Chomsky 1981, 1986, Chomsky & Lasnik 1993] and minimalism on the other hand [Chomsky 1995, 2000, in press]. Actually I think of myself as rather eclectic in this perspective, I just pick stuff left and right, which I happen to think is our best bet at being right. Another thing which actually predates the minimalist discussions is the status of parameters. I mean, at some point it became fashionable to say: "Well, parameters are sort of designed to do language-particular stuff" — but especially as more micro-syntactic stuff came up [see Kayne 2000 for discussion and further references, for example], it became clear that these big, overall sweeping parameters didn't really work that way, and that they were much more fine-grained and so on.

So it didn't seem a bad idea to say that if that's sort of language-particular stuff, why don't we put it where the language-particular stuff belongs, namely in the lexicon. Ok, fine, but the other thing that was always said about it was that, we know that people learn words the hard way — sort of, you know, they have to pick them up from experience — and parameters clearly have to be learned because they need to be set, so that's a good reason why we should put them in the lexicon. And as for parameters that are not tied to the lexicon — I mean, despite things like the subset principle and so on, it's not entirely clear how they are learned.

Probably biological.

Well, frankly, that fails to convince me. I mean, the primary data that are needed to fix something like the Headedness Parameter, is perfectly straightforward, and I have always failed to see the rationale for trying to get rid of the Headedness Parameter. Maybe if it's possible for us to learn that the English word *cheese* is a noun that means what is does, then why can't we learn some little property that says, "In this language heads are final"? Big deal. You know, maybe it's like a big silent word that sits in the lexicon and that we happen not to use to insert into syntactic structures, but that we use as a reference library. I don't see anything wrong with that in principle. And, of course, that gets us into the whole LCA business [Kayne 1994] and so on. If we look at what a fairly simple thing like the Headedness Parameter buys us and all the trouble that it avoids for us, well, I think I'll settle for that any time.

Sounds good. Here's a potential shortcoming of the minimalist approach in this respect — any minimalist approach, from the beginning within the P&P theory — that might be formulated: On the P&P approach — you just mentioned it, that when it was developed, we had some core principles, whatever they are, and it's true that, if it's not exactly enough to have the big parameters, people will look for smaller ones. If it's this type of approach looking at the specification of these parameters — take Clark & Roberts's [1993] work on parameters, for example — then we haven't really made that much progress in terms of how we should visualize these purported principles and parameters. Along comes minimalism, which was quite a different approach — without explicitly saying, "Ah, principles and parameters." No, it says that we can build in parametric variation — strong features, weak features. But does that really tell us anything deeper? How does that fit in — given that Chomsky and others, you and me included, spent a lot of time on architecture, architecture and design. But what is the architecture and design of these principles and parameters? Is it the switch-board analogy [which Chomsky attributes to Jim Higginbotham in this volume; Chomsky 2007]? Is it the Baker-style way, that you can go to parameter 3 only if parameter 2 is switched on [cf. Baker 2001]? If parameter 2 *is switched off, then you never get to parameter 3, so it's irrelevant for, let's say, 95% of the languages. I don't know of too much literature that deals exactly with these kinds of questions — apart from the learnability angle, Clark & Roberts and others — I don't know of too much work that tells us something more about design and architecture of principles, and of parameters, regardless of what they are. And I don't think people have really tackled that question in the last 15 years.*

Well, I think you are exactly right. Basically, I don't think that in the P&P period people were so lazy. I mean, lots of people were working on parameters, and one of the main things that was happening was that with increased attention for dialects and for more related languages, that is, with increased interest for micro-syntax, people started to realize that these parameters really are about much more fine-grained structure. Well, that was about where we arrived at, when minimalism, you know, hit us around the ears, and that put an end to that. That is, thankfully, research on micro-syntax is still going on, but you don't see these people talking about parameters much anymore.

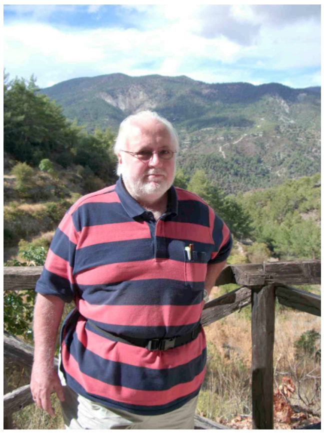
Do you think they would have continued? You could say, "Oh no, minimalism came at the right time" — because at the time, we were not ready yet to spell out our assumptions about the structure of principles and parameters clearer. Bear in mind that all I just said, that was never spelled out by anyone, of course, and Chomsky didn't say, "Hey, let's do this new minimalism thing now because we are not ready yet to continue with the old line of research on principles and parameters."

Yes, I think it's likely that someone would have continued talking about parameters.

So, is the future perhaps Principles–and–Parameters and *biolinguistics?*

We'll see.

Thank you very much for a stimulating conversation and a wonderful day.



References

Atkinson, Martin. 2007. Mother Merge and her children. Ms., University of Essex, Colchester.

Baker, Mark. 2001. *The Atoms of Language: The Mind's Hidden Rules of Grammar*. New York: Basic Books.

Boeckx, Cedric & Kleanthes K. Grohmann. 2007. The *Biolinguistics* manifesto. *Biolinguistics* 1, 1-8.

Breitbarth, Anne & Henk van Riemsdijk (eds.). 2004. *Triggers* (Studies in Generative Grammar 75). Berlin: Mouton de Gruyter.

Carnie, Andrew & David Medeiros. 2005. Tree maximization and the generalized Extended Projection Principle. *Coyote Working Papers in Linguistics* 14, 51-55.

Carnie, Andrew, David Medeiros & Cedric Boeckx. 2005. Some consequences of natural law in syntactic structures. Ms., University of Arizona, Tucson & Harvard University, Cambridge.

Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms., Harvard University/Massachusetts Institute of Technology. [Published in part as *The Logical Structure of Linguistic Theory*, New York: Plenum, 1975.]

Chomsky, Noam. 1956. Three models for the description of language. *IRE Transactions on Information Theory* 2(3), 113-124.

Chomsky, Noam. 1957. *Syntactic Structures* (Januar Linguarum, Series Minor 4). The Hague: Mouton.

Chomsky, Noam. 1959. Review of B.F. Skinner, *Verbal Behavior*. *Language* 35(1), 26-58. [Reprinted in Barbara C. Lust & Claire Foley, eds. 2004. *First Language Acquisition: The Essential Readings* (Linguistics: The Essential Readings), 25-55. Malden, MA: Blackwell.]

Chomsky, Noam. 1965. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press.

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

Chomsky, Noam. 1982. *Some Concepts and Consequences of the Theory of Government and Binding* (Linguistic Inquiry Monograph 6). Cambridge, MA: MIT Press.

Chomsky, Noam. 1986. *Barriers* (Linguistic Inquiry Monograph 13). Cambridge, MA: MIT Press.

Chomsky, Noam. 1994. Bare phrase structure. *MIT Ocassional Papers in Linguistics* 5. [Reprinted in Webelhuth, Gert (ed.) 1995. *Government and Binding Theory and the Minimalist Program* (Generative Syntax 1). Malden, MA: Blackwell, 383-439.]

Chomsky, Noam. 1995. *The Minimalist Program* (Current Studies in Linguistics 28). Cambridge, MA: MIT Press.

Chomsky, Noam. 2000. Minimalist inquiries: The framework. In Roger Martin, David Michaels & Juan Uriagereka (eds.), *Step by Step: Essays on 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* (Current Studies in Linguistics 36), 1-52. Cambridge, MA: MIT Press.

Chomsky, Noam. 2004. Beyond explanatory adequacy. In Adriana Belletti (ed.), *Structures and Beyond*, vol. 3: *The Cartography of Syntactic Structures* (Oxford Studies in Comparative Syntax), 104-131. New York: Oxford University Press.

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

Chomsky, Noam. 2007. Of minds and language. *Biolinguistics* 1, 9-27.

Chomsky, Noam. In press. On phases. In Freidin et al. (eds.).

Chomsky, Noam & Howard Lasnik. 1993. The theory of principles and parameters. In Joachim Jacobs, Arnim von Stechow & Wolfgang Sternefeld (eds.) *Syntax: Ein internationales Handbuch zeitgenössischer Forschung/An International Handbook of Contemporary Research*, vol. 1 (Handbücher zur Sprach- und Kommunikationswissenschaft/Handbooks of Linguistics and Communication Science 9/1), 506-569. Berlin: Mouton de Gruyter. [Reprinted in Chomsky (1995), 13-127.]

Clark, Robin & Ian Roberts. 1993. A computational model of language learnability and language change. *Linguistic Inquiry* 24(2), 299-345.

Freidin, Robert, Carlos P. Otero & Maria Luisa Zubizarreta (eds.). In press. *Foundational Issues in Linguistic Theory: Essays in Honor of Jean-Roger Vergnaud* (Current Studies in Linguistics). Cambridge, MA: MIT Press

Grohmann, Kleanthes K. 2000. Prolific peripheries: A radical view from the left. College Park, MD: University of Maryland dissertation.

Grohmann, Kleanthes K. 2001. "Natural relations": A note on X'-structure. *ZAS Papers in Linguistics* 21, 67-87.

Grohmann, Kleanthes K. 2003. *Prolific Domains: On the Anti-Locality of Movement Dependencies* (Linguistik Aktuell/Linguistics Today 66). Amsterdam: John Benjamins.

Hornstein, Norbert. 2001. *Move! A Minimalist Theory of Construal* (Generative Syntax 5). Malden, MA: Blackwell.

Hornstein, Norbert, Jairo Nunes & Kleanthes K. Grohmann. 2005. *Understanding Minimalism* (Cambridge Textbooks in Linguistics). Cambridge: Cambridge University Press.

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

Jenkins, Lyle (ed.). 2004. *Variation and Universals in Biolinguistics* (North-Holland Linguistics Series: Linguistic Variations 62). Amsterdam: Elsevier.

Kayne, Richard S. 1994. *The Antisymmetry of Syntax* (Linguistic Inquiry Monograph 25). Cambridge, MA: MIT Press.

Kayne, Richard S. 2000. *Parameters and Universals* (Oxford Studies in Comparative Syntax). New York: Oxford University Press.

Leonard, Larry. 2004. Exploring the phenotype of specific language impairment: A look at grammatical variability. In Jenkins (ed.), 285-316.

Medeiros, David. 2006. A case of ontological minimalism: X-bar as an optimal packing solution. Ms., University of Arizona, Tucson.

Piattelli-Palmarini, Massimo. 1974. A debate on bio-linguistics. *Centre Royaumont pour une science de l'homme report*. [Conference held at Endicott House, Dedham, Massachusetts, 20-21 May, 1974.]

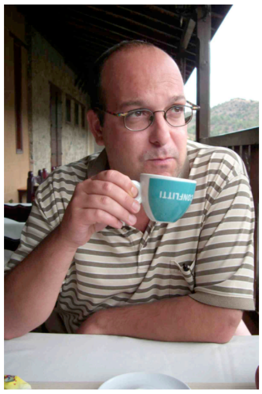
Piattelli-Palmarini, Massimo (ed.). 1980. *Language and Learning: The Debate between Jean Piaget and Noam Chomsky*. London: Routledge & Kegan Paul.

Piattelli-Palmarini, Massimo & Juan Uriagereka. 2004. The immune syntax: The evolution of the language virus. In Jenkins (ed.). 342-378.

Piattelli-Palmarini, Massimo, Juan Uriagereka & Pello Salaburu (eds.). In press. *Of Minds and Language: The Basque Country Encounter with Noam Chomsky*. New York: Oxford University Press.

van Riemsdijk, Henk. 2001. A far from simple matter: Syntactic reflexes of syntax-pragmatics misalignments. In István Kenesei & Robert M. Harnish (eds.),

Perspectives on Semantics, Pragmatics, and Discourse: A Festschrift for Ferenc Kiefer (Pragmatics & Beyond New Series 90), 21-41. Amsterdam: John Benjamins.
van Riemsdijk, Henk. 2006a. Free relatives. In Martin Everaert & Henk van Riemsdijk (eds.), *The Blackwell Companion to Syntax*, vol. II, 338-382. Oxford: Blackwell.
van Riemsdijk, Henk. 2006b. Grafts follow from merge. In Mara Frascarelli (ed.) *Phases of Interpretation* (Studies in Generative Grammar 91), 17-44. Berlin: Mouton de Gruyter.
van Riemsdijk, Henk. 2006c. Organizing grammar. Invited talk at the *Department of Humanities Lecture Series*, Nicosia. [Cyprus College, 11 October 2006.]
van Riemsdijk, Henk. 2006d. External and internal merge. Invited lecture in the Department of Humanities, Nicosia. [Cyprus College, 12 October 2006.]
van Riemsdijk, Henk. In press. Identity avoidance: OCP-effects in Swiss relatives. In Freidin *et al.* (eds.).
Soschen, Alona. 2007. On the nature of syntax. Ms., MIT, Cambridge.
Uriagereka, Juan. 1998. *Rhyme and Reason: An Introduction to Minimalist Syntax*. Cambridge, MA: MIT Press.
Uriagereka, Juan. 1999. In defense of deep structure. Ms., UNiversity of Maryland, College Park. [To appear as chapter 1 of Uriagereka, Juan. In press. *Syntactic Anchors: On Semantic Structuring* (Cambridge Studies in Linguistics 118). Cambridge: Cambridge University Press.]
Vergnaud, Jean-Roger. 1977. Personal letter to Howard Lasnik and Noam Chomsky, Paris. [To be reprinted in Freidin *et al.* (in press).]
Wexler, Ken. 2004. Lenneberg's dream: Learning, normal language development and specific language impairment. In Jenkins (ed.), 239-284.



Henk van Riemsdijk
Tilburg University
Faculty of Arts
P.O. Box 90153
5000 LE Tilburg
The Netherlands
riemsdyk@uvt.nl

Kleanthes K. Grohmann
University of Cyprus
Department of English Studies
P.O. Box 20537
1087 Nicosia
Cyprus
kleanthi@ucy.ac.cy

★ FORUM ★

Notice

We would like to use this opportunity to thank all those involved in creating the first volume of *Biolinguistics* and all the work that went into preparing the entire endeavour over the past year and a half or so. For internal quality control, our gratitude goes to the reviewers that have served us in 2007, who are listed below. For everything else, we thank our supporters listed below as well as all active participation and feedback from the members of the *Biolinguistics* Advisory Board, the *Biolinguistics* Editorial Board, and the *Biolinguistics* Task Team that are not specifically mentioned by name.

Reviewers

Mark C. Baker
Željko Bošković
Carlo Cecchetto
John Drury
Maria-Teresa Guasti
Wolfram Hinzen
Jason Merchant
Andrew Nevins
Jairo Nunes
Dennis Ott
Phoevos Panagiotidis
Anna Roussou
Bridget Samuels
Volker Struckmeier
Ianthi Maria Tsimpli
Juan Uriagereka
Masaya Yoshida
Jochen Zeller

Supporters

Stephen R. Anderson
Noam Chomsky
John Davey
Marcel den Dikken
Anna Maria diSciullo
Samuel David Epstein
Norbert Hornstein
So-One Hwang
Lyle Jenkins
David Lightfoot
Terje Lohndal
Frédéric Mailhot
Jeffrey K. Parrott
Massimo Piattelli-Palmarini
Pierre Pica
Henk van Riemsdijk
Marinos Stylianou
George Tsoulas
Constantinos Tsouris
Juan Uriagereka
Andreas Wadle
Andrew Winnard

Biolinguistics
Volume 1
Fall 2007

TABLE OF CONTENTS

★ **EDITORIAL** ★

001	The *Biolinguistics* Manifesto	Cedric Boeckx *Harvard University* Kleanthes K. Grohmann *University of Cyprus*
009	Of Minds and Language	Noam Chomsky *MIT*
028	Computing Long-Distance Dependencies in Vowel Harmony	Frédéric Mailhot *Carleton University* Charles Reiss *Concordia University*
049	Categorical Acquisition: Parameter Setting in Universal Grammar	Rosalind Thornton *Macquarie University* Graciela Tesan *Macquarie University*
099	Clarifying the Notion "Parameter"	Juan Uriagereka *University of Maryland*

★ **BRIEFS** ★

114	A Note on the Default Values of Parameters	Koji Sugisaki *Mie University*

★ **REVIEWS** ★

118	Juliette Blevins, *Evolutionary Phonology* (CUP, 2004)	Bridget Samuels *Harvard University*

★ **FORUM** ★

123	Continuous Acceptability, Categorical Grammaticality, and Experimental Syntax	Jon Sprouse *University of California, Irvine*
135	The Anatomy of Biolinguistic Minimalism	Samuel David Epstein *University of Michigan* T. Daniel Seely *Eastern Michigan University*
137	Interview with Henk van Riemsdijk	Kleanthes K. Grohmann *University of Cyprus*
150	Notice	*Biolinguistics* Editors

biolinguistics

ISSN 1450–3417

Biolinguistics 1: 001–150, 2007
http://www.biolinguistics.eu