50 YEARS LATER
REFLECTIONS ON CHOMSKY’S ASPECTS

edited by Ángel J. Gallego and Dennis Ott

2015

The printed version of this volume is available from the MITWPL online store:
http://mitwpl.mit.edu

The copyright remains with the individual authors.
50 YEARS LATER
REFLECTIONS ON CHOMSKY’S ASPECTS

CONTENTS

EDITORS’ INTRODUCTION
Ángel J. Gallego and Dennis Ott................................................................. iii

FOREWORD
David Pesetsky............................................................................................... v

PREFACE
Noam Chomsky............................................................................................... vii

MIND THE GAP
Robert C. Berwick......................................................................................... 1

QUESTIONS WITH DECLARATIVE SYNTAX TELL US WHAT ABOUT SELECTION?
Jonathan David Bobaljik and Susi Wurmbrand........................................... 13

THE SUBTLE DEPENDENCY BETWEEN COMPETENCE AND PERFORMANCE
Cristiano Chesi and Andrea Moro................................................................. 33

A SOURCE OF PARAMETRIC VARIATION IN THE LEXICON
Guglielmo Cinque.......................................................................................... 47

RELATIVE CLAUSE DELETION
Chris Collins................................................................................................. 57

3RD YEAR GRAMMAR
Stephen Crain and Rosalind Thornton....................................................... 71

RAISING THE SUBJECT OF THE ‘OBJECT-OF’ RELATION
Marcel den Dikken....................................................................................... 85

FROM ASPECTS’ ‘DAUGHTERLESS MOTHERS’ TO POP’S ‘MOTHERLESS SETS’: A
SELECTIVE HISTORY OF THE EVOLUTION OF SIMPLEST MERGE
Samuel D. Epstein, Hisatsugu Kitahara and T. Daniel Seely...................... 99

ON THE ALLEGED ABSENCE OF CERTAIN AMBIGUITIES IN FREE CONSTITUENT ORDER
LANGUAGES
Gisbert Fanselow......................................................................................... 113

A NOTE ON WEAK VS. STRONG GENERATION IN HUMAN LANGUAGE
Naoki Fukui.................................................................................................. 125

A NOTE ON ENGLISH SUBJECT CONTACT RELATIVES
Liliane Haegeman......................................................................................... 133
<table>
<thead>
<tr>
<th>Title</th>
<th>Author(s)</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>THE RATIONALISM OF GENERATIVE GRAMMAR</td>
<td>Norbert Hornstein</td>
<td>147</td>
</tr>
<tr>
<td>ASPECTS AND BEYOND: THE CASE FOR GENERALIZED CONSTRUAL</td>
<td>Jan Koster</td>
<td>159</td>
</tr>
<tr>
<td>ASPECTS OF THE THEORY OF PHRASE STRUCTURE</td>
<td>Howard Lasnik</td>
<td>169</td>
</tr>
<tr>
<td>HOW TO TRIGGER ELEMENTS OF I-LANGUAGES</td>
<td>David W. Lightfoot</td>
<td>175</td>
</tr>
<tr>
<td>THE GARDEN OF EDEN PERIOD FOR DEEP STRUCTURE AND SEMANTICS</td>
<td>Barbara H. Partee</td>
<td>187</td>
</tr>
<tr>
<td>VOCABULARY MATTERS</td>
<td>Paul M. Pietroski</td>
<td>199</td>
</tr>
<tr>
<td>LEIBNIZIAN LINGUISTICS</td>
<td>Ian Roberts and Jeffrey Watumull</td>
<td>211</td>
</tr>
<tr>
<td>NOMINAL DECAY</td>
<td>Haj Ross</td>
<td>223</td>
</tr>
<tr>
<td>FEATURES FROM ASPECTS VIA THE MINIMALIST PROGRAM TO COMBINATORY</td>
<td>Neil Smith and Annabel Cormack</td>
<td>233</td>
</tr>
<tr>
<td>CATEGORIAL GRAMMAR</td>
<td></td>
<td></td>
</tr>
<tr>
<td>METALISM AND UNIVERSALS IN ASPECTS</td>
<td>Tim Stowell</td>
<td>249</td>
</tr>
<tr>
<td>ON BY-PASSIVE AND WHY UNACCUSATIVES DON’T PASSIVIZE</td>
<td>Tarald Taraldsen</td>
<td>255</td>
</tr>
<tr>
<td>COMPLEMENTARITY IN LANGUAGE</td>
<td>Juan Uriagereka</td>
<td>265</td>
</tr>
<tr>
<td>GENERALIZED TRANSFORMATIONS REANIMATED</td>
<td>Edwin Williams</td>
<td>279</td>
</tr>
<tr>
<td>INPUT AND ITS STRUCTURAL DESCRIPTION</td>
<td>Charles Yang, Allison Ellman, and Julie Anne Legate</td>
<td>285</td>
</tr>
</tbody>
</table>
It is customary to refer to Noam Chomsky’s *Aspects of the Theory of Syntax* as the founding document of Generative Grammar—and justifiably so. It was not Chomsky’s first book, of course: his *The Logical Structure of Linguistic Theory* and *Syntactic Structures* had been completed almost a decade earlier. But the former only circulated in mimeographed form until it was finally partially published in 1975; and the latter was a rather informal write-up of lecture notes appearing with a small publisher in the Netherlands (as was the case with Chomsky’s monograph *Current Issues in Linguistic Theory*, a year ahead of *Aspects*).

*Aspects*, in other words, was Chomsky’s first thoroughly worked-out book-length articulation of the generative program and its significance to other fields, widely received thanks to the backing of a renowned publisher. Its groundbreaking first chapter on “Methodological Preliminaries” lucidly carved out the fundamentals of the nascent approach, defining the object of inquiry as “the ideal speaker’s intrinsic competence” (dubbed *I*-language in Chomsky’s later works) and thereby introducing the competence/performance distinction. It is not too far-fetched to say that this was the move that placed linguistics on scientific footing, by shifting the focus from the common-sense notion of “language” to a well-defined domain of inquiry.

Paving the way for Chomsky’s philosophical follow-up *Cartesian Linguistics* (1966), this first part of *Aspects* is also an impressive defense of the rationalist view of language acquisition/growth—only on this conceptual foundation, Chomsky reasoned, can we feasibly aspire to *explanatory* (rather than merely observational or descriptive) adequacy. The rest is history: the quest for explanatory adequacy as defined in *Aspects* fueled the crystallization of the Principles & Parameters approach; and more recently, Chomsky launched the Minimalist Program as an attempt to go beyond this goal. Tempting as it is, we refrain at this point from going further into the profound impact this first part of *Aspects* had on the field; its lasting influence is aptly summarized in David Pesetsky’s foreword to this volume.

The second part of *Aspects* moves beyond the conceptual level and develops certain technical details of the generative procedure in some detail. For the first time, *Aspects* presented a formal
model of grammar comprising transformationally related levels of representation, now fed by a lexicon. Some of the technical details, such as the use of features in subcategorization frames or the matching analysis of relativization, continue to figure prominently in the literature. Many others have, of course, been revised or replaced since—for instance, Aspects’ shift away from Generalized Transformations in favor of a recursive base component and Chomsky’s preference for concatenation over set formation have seen dramatic reversals in recent work, Chomsky’s own in particular.

This volume is a celebration of Aspects’ multifaceted impact on the field of linguistics, and a reflection on some of its leading themes 50 years into the game. As the reader will no doubt notice, the contributions in this volume differ quite significantly, ranging from short remarks to detailed technical analyses. We, the editors, deliberately chose not to counteract this heterogeneity, which bears testimony to the wide range of ideas and visions that Aspects continues to inspire. We hope that the volume persuades its reader to revisit Aspects, which we believe is a worthwhile exercise not only for historical reasons—after all, many of its central theses concerning the biological nature of language continue to be subject of debate, as Noam Chomsky points out in his preface to this volume. It’s sometimes said that if you lose your way in the forest (a sensation that some of us may have nowadays in view of an increasingly fragmented field), it can be helpful to return to your starting point, from where you may be able to discern your original direction. In much the same vein, we believe, Aspects can serve as a reminder of where we were headed back in 1965, and still are today.
FOREWORD*

DAVID PESETSKY
Massachusetts Institute of Technology

Rereading Noam Chomsky’s *Aspects of the Theory of Syntax* almost a half-century after its 1965 publication, I found myself equally startled by two contradictory observations: how breathtakingly modern this book remains, and how old-fashioned it has become.

*Aspects* was actually the second book by Chomsky to revolutionize the study of language. It was preceded in 1957 by *Syntactic Structures*, which showed how a small set of formal tools could provide a simple model for astonishingly complicated facts about English syntax (most of which had gone unnoticed until then). Thanks to a glowing review in the journal *Language*, *Syntactic Structures* was widely read. Overnight, Chomsky’s “transformational grammar” became an area of intensive research. The pace of discovery was extraordinarily rapid. Paper after paper presented brand-new discoveries about syntax, prompting many hotly debated revisions of the basic model.

But what was this new field actually all about? Why should we care about the best description of a group of verbs, or the position of an adjective? It was *Aspects* that offered some answers—so compelling and incisive that they continue to ground our research to this very day. *Aspects* set the following goals for linguistics:

1. Language makes “makes infinite use of finite means” (a phrase from Humboldt that *Aspects* made famous). The right theory of language “must describe the processes that make this possible.”
2. The minute one examines the facts of language, one is struck by the incredible complexity of linguistic phenomena—yet every normally developing child masters this complexity at a very young age, effortlessly. Compounding the puzzle is the inescapable fact of linguistic diversity. Languages differ enormously (different vocabulary, different word orders, a different array of constructions) yet any child can acquire any one of them. The right theory of language must therefore be “sufficiently rich to account for acquisition of language, yet not so rich as to be inconsistent with the known diversity of language.”

* This text was first published on MIT Press’ online blog (https://mitpress.mit.edu/blog/1965-aspects-theory-syntax) and is reprinted here with kind permission of MIT Press.
3. The devil is in the details. The right theory of language must be “perfectly explicit”—what Chomsky called a “generative grammar.” As in any science, linguistic hypotheses must predict, and therefore explain, the facts.

These are the goals that most of us in linguistics pursue to this very day, hence the breathtakingly modern feel of this great book. At the same time, although *Aspects* was the first work to articulate the central tension between the ease of language acquisition (explainable if much of language is innate) and linguistic diversity (explainable if much of language is not innate), and though Chomsky spent many pages “eating his own dog food,” by working out the details—I am also struck by how many of these details have been superseded and replaced by newer, better ideas.

But this is actually the most important sign of *success*, not failure, for this great book. The fact is, in 1965 very little was known about how children actually acquire language, and almost nothing was known about the extent and limitations of cross-linguistic diversity in syntax. In the wake of *Aspects*, the last half-century of linguistics has been marked by an enormous explosion of knowledge in both domains. And just as *Aspects* teaches us, this means that the details of our “perfectly explicit” theory had to change to match our new knowledge—with Chomsky himself in the forefront of many of these developments. The greatest works in science contain a timeless core, but if they are worth anything at all, they also contain the seeds of their own obsolescence—because the greatest works are those that launch a living field, and living fields do not stand still. So it is with *Aspects*.
Aspects of the Theory of Syntax was the first attempt to present a general framework for the work in generative grammar that had been developing since the late 1940s, with applications in a number of languages (in rough chronological order, Hebrew, English, German, Russian, Hidatsa, Turkish, Mohawk, Japanese, and some others, at varying levels of depth). It also discusses a variety of other problems of language, most of them still alive, even if in different forms. It was also one of several publications that sought to place these efforts within a larger context of earlier thinking about language and the mind that had been mostly forgotten or seriously misrepresented.¹

This work was based on a number of assumptions that seem elementary but that departed from views that were prevalent at the time in the related disciplines. The first of these assumptions is that each language is a biological property of the individual language user, mostly parts of the brain. The biological framework was greatly enriched by Eric Lenneberg’s seminal work (1967). A few years later the approach began to be called “the biolinguistic program,” adopting a term suggested by Massimo Piattelli-Palmerini, who organized a major conference on these topics in 1974 bringing together biologists, linguists, philosophers and psychologists, held at MIT in conjunction with the Royaumont Center in Paris.

The second basic assumption had to do with the nature of language. Each language makes available an unbounded array of hierarchically structured expressions that have determinate interpretations at the interfaces with other internal systems: systems of thought and organization of action (the conceptual-intentional interface CI), and the sensorymotor system SM for externalization (production and perception); usually sound, though as is now known, other modalities are possible. We may call this core feature of language its Basic Principle.

The Basic Principle comprehends the entire computational aspect of language, syntax in the broad sense, including the narrow syntax that provides the expressions mapped to the interfaces and the mappings themselves, and of course the lexical atoms of computation and their various configurations.

At the time, each individual language viewed from this perspective was called “a grammar,” in one of the uses of this systematically ambiguous expression. Adopting a later terminological suggestion, the system is an I-language, where “I” signifies individual, internal, and intensional (in that we are interested in the actual generative procedure, not some set of entities that it determines: a set of utterances, structures, etc.). The theory of an I-language is a (generative) grammar.

Languages can vary within the limits set by the genetic factors that specify the language faculty, called “universal grammar” (UG) in contemporary terminology. The general theory of language seeks to give an explicit account of UG, thus identifying the true nature of the Basic Principle.

Several goals for linguistic theory are discussed in Aspects. For particular grammars, the basic goal is essentially truth: to provide a correct theory of the I-language under consideration. To the extent that it does, the grammar meets the condition of descriptive adequacy. This is, needless to say, an extraordinarily difficult task, and even approximating it for a single language would be quite an achievement. We can regard a linguistic theory as descriptively adequate to the extent that it provides descriptively adequate grammars for every human I-language, real or attainable, and no others, thus determining exactly how the Basic Principle can be instantiated in human language. That is an even more far-reaching goal. But there is a deeper problem to be confronted as well: what is called explanatory adequacy in Aspects. A linguistic theory satisfies this condition to the extent that it shows how a descriptively adequate grammar for each I-language is selected on the basis of the data available (primary linguistic data), and thus shows how in principle languages can be learned, one crucial part of the biolinguistic program. From this point of view, the task of selection of a grammar is what came to be called “the logical problem of language acquisition.”

As in every domain of science, we seek the simplest theory, the one with the most far-reaching capacity for explaining phenomena, rather than just describing them. But within the biolinguistic program, there is an additional reason to seek the simplest theory of UG. It must have evolved, and the more complex it is, the greater the obstacles to some eventual theory of the origin of language – and the word eventual should be stressed, though there are a few useful hints.

As soon as the first careful efforts were undertaken to construct explicit grammars for particular languages, it was discovered that very little was understood, contrary to widespread belief. Major problems appeared everywhere, many of them never noticed in the thousands of years of intensive and productive research into language. It was also quickly discovered that from the earliest stages of acquisition, language resembles other biological systems in that what is attained is vastly underdetermined by the evidence available. This is understood to be the merest truism in other domains of biology, and is therefore not dignified by a name. In the study of language, even what seem to be truisms are highly contested, so there is a name: Poverty of Stimulus (POS). Many illustrations are given in Aspects and earlier publications. Most have been ignored by those who contest the significance of POS for language. A few have been addressed, mostly rather marginal questions, but with very little success, to my knowledge.

In accord with normal scientific practice, the study of an I-language should seek to sort out the many variables that enter into what is directly observed in order to focus on what is relevant

---

2 Surprisingly, this scientific truism is sometimes denied (Margolioash and Nusbaum, 2009), but on the basis of serious misunderstanding.

3 One significant case has received considerable attention, structure-dependence of rules. Berwick et al. (2011) reviews every proposal that was clear enough to examine. All fail, irremediably, and in fact are asking the wrong question. More recent proposals, to my knowledge, suffer the same fate. There is a simple and well-confirmed answer, with interesting consequences, but it is never considered in this general literature. See Chomsky (2013) for a non-technical review.
for the purpose at hand. In the terminology of *Aspects*, research seeks to investigate an “ideal speaker-hearer,” putting to the side non-linguistic factors that enter actual performance (memory constraints, lapses of attention, etc.). A related notion is that inquiry should distinguish competence (roughly, what the person knows) from performance (what the person does in particular circumstances).

All of this should be – in fact is – second nature to every researcher. Thus in field work with an informant, the first step is transcription of utterances, which already involves high-level idealizations, disregarding non-linguistic factors and adopting non-trivial assumptions about the linguistic factors, thus abstracting to a postulated ideal speaker-hearer and distinguishing competence from performance. Any further inquiry simply carries the process further. The same holds for investigation of any biological system, in fact anything in the empirical world. Every experiment, for example, is based on essentially these distinctions, seeking to extricate the matter of interest from what is background noise for the purpose of the inquiry.

An elementary illustration in the domain of syntax is the widely discussed property of embedding – often confused with recursion. English permits such sentences as (1) but not (2):

1. a. If it’s hot then Bill will be happy
   b. Either it’s hot or it’s winter
   c. Days that Bill likes are rare

2. a. If it’s hot or Bill will be happy
   b. Either it’s hot then it’s winter
   c. Days that Bill likes is rare

Furthermore, these can be embedded in one another, as in (3), where (1b) is embedded in (1a):

3. If either it’s hot or it’s winter then Bill will be happy

Such embedding yields nested dependencies (if-then, either-or, plural-plural, etc.), and it is easy to show that unbounded nested dependencies cannot be accommodated by any of the theories of language proposed in the 1950s, the main reason why these structures received attention at the time.⁴

In 1963, George Miller and I observed that without external memory, people can understand repeated embedding up to about depth six. That is as expected, in the light of the general limits of short-term memory across species.⁵ With external memory and time, of course, the limits can be exceeded, in principle indefinitely, much as in the case of arithmetical operations.

The rules of the I-language in the case of (1)-(3) are transparent. Plainly, it would be absurd to add to them a limit that happens to be the limit of short-memory, with some extra statement about how it can go on without limit when memory and other language-independent constraints are lifted.

Restating the observations in the terminology of *Aspects*, unbounded self-embedding (and nesting of dependencies) is part of the speaker’s competence, but performance is constrained by language-independent memory factors.

---

⁴ Another reason was that they fall within varieties of phrase structure grammar (PSG), though as the relevant publications of the time emphasized, all varieties of PSG fail on the crucial scientific grounds of explanatory adequacy, an observation summarized and extended in *Aspects*.

All of this should be obvious. Nevertheless, these notions were highly contested at the time, and since.

The most far-reaching objection is to deny the Basic Principle, by now a very fashionable position, hence perhaps worth some attention. Perhaps the first explicit challenge was in a highly influential paper of Rumelhart and McClelland (1986), which introduced connectionist/neural net models of language. They deny that productivity is “of the essence of human computation” on the grounds that center-embedded sentences are hard to process – as indeed they are, just as large numbers are hard to multiply in one’s head. The argument is by now common. For example, Margoliash and Nusbaum (2009) write that “In reality, humans rarely achieve, and only awkwardly, even a recursive level of three when using center embedded sentences in natural speech.” They therefore propose that “embracing an infinite recursion from linguistics unsupported by corresponding data of actual human performance arises from an unfettered first-principle perspective” of the kind they condemn as ignoring the modern biological sciences. To take another case, Stephen Levinson (2013) reaches the same conclusion on the basis of the observation that center-embedding of even degree two is “vanishingly rare” in actual corpora, and degree three “almost never” appears.

The observations are correct, and explicitly refute the conclusions drawn from them. It is entirely true that for center-embedding beyond degree two, data are “vanishingly rare,” in fact virtually non-existent. Nevertheless, as noted earlier, it has been known for half a century that people can process up to the limits of short-term memory (6 or 7) with no external aids. It follows at once that without exposure to relevant data, children know the obvious rules and in particular that they have no bounds. It follows further that that the models proposed by those who reject the Basic Principle are radically incorrect for natural language. The conclusion is indeed based on “an unfettered first principle perspective,” namely the principles of rationality taken for granted in the modern biological sciences as elsewhere.

The analogy to arithmetic, just mentioned, brings out the essential point in an elementary and uncontroversial way. The arithmetical competence that appears to be a common human possession provides triples (x, y, z) where z is the product of x and y, but without external aids (memory, time, attention, life span,…) multiplication of large numbers is “vanishingly rare.” The same is true of everyone’s laptop. It’s possible to provide it with a calculator, but without access to external memory, it will have bounded performance. Lacking any such access, the system has to be redesigned for larger calculations. With such access, it needs no redesign. That is the familiar distinction between strictly finite automata and Turing architecture.

All of this is elementary. It suffices to show that the favorite example of those who reject the Basic Principle and the competence-performance distinction refutes their assumptions, and all that goes with them.

At least some of the objections to the general framework outlined in Aspects make sense on the assumption that there is nothing special about human language: it is just “more of the same.” In some versions, “the same” is human cognitive processes, in which case the properties of language reduce to the analogue of UG for these processes generally. In other versions, “the same” holds more broadly for non-human organisms as well. Under these assumptions, POS of course remains, as it must, but it is not specific to language.

Prima facie, this stand seems to be rather hard to sustain, if only because human language is so radically different from anything else observed in the natural world. Many more specific reasons were brought out in Lenneberg’s work on biology of language, in particular, the (double)
dissociations between language capacity and other cognitive faculties. By now there is rich data on these dissociations.\textsuperscript{6} Such facts seem hard to reconcile with the “nothing special” assumptions.

Another reason is that a new-born infant at once selects language-relevant data from the “blooming buzzing confusion” with which it is confronted. Certain brain areas selectively react to human speech rather than other noises, becoming more specialized during the first few months of life.\textsuperscript{7} After that, language development typically follows a regular course, on to full mature competence. In contrast, an ape with about the same auditory apparatus (or visual apparatus in the case of sign), does not take the first step, let alone those that follow. Again, it seems fairly evident that “something special” is involved.

The crucial questions however have to do with explanation of the properties of language that have been discovered. This is, of course, not the place to review that the status of these considerations. I will only express my personal view that the “more of the same” conception, despite its popularity, has very little to recommend it and is largely a recipe for continued failure to learn about the nature, use, and acquisition of language.

With some qualifications, the conclusion about these matters in \textit{Aspects} seems to me to remain valid: “there is surely no reason today for taking seriously a position that attributes a complex human achievement entirely to months (or at most years) of experience rather than to millions of years of evolution or to principles of neural organization that may be even more deeply grounded in physical law.” The same holds if we replace “entirely” by “mostly.” The major qualifications have to do with the length of the relevant evolutionary process. There is by now considerable evidence that the relevant time frame may be very brief and recent. And there is increasing evidence that general principles of computational complexity, perhaps grounded in physical law, are of fundamental significance in acquisition of I-language.

As I mentioned, \textit{Aspects} was a first attempt to develop a framework for theories of language that accommodated the Basic Principle within an emerging biolinguistic framework. While the general approach seems to me essentially correct, there have, of course, been very substantial developments since.

One crucial move was to revise the \textit{Aspects} approach to selection of grammar given data. The format-and-evaluation approach of \textit{Aspects}, which derived from work of the late 1940s, suffered from a serious defect, as recognized there: unfeasibility. In principle, it provided an answer, but only with astronomical amounts of computation. A solution to the problem was offered by the Principles and Parameters framework that crystallized in the early 1980s, opening what proved to be a highly successful research program. It revitalized the study of language acquisition, which has made impressive progress since. It opened the way to a vast explosion in range and depth of descriptive work on languages of great typological variety. There has also been very illuminating work on the structure and variety of parameters, with consequences for comparative and historical linguistics. It also made possible a deeper inquiry that seeks to reduce the architecture of I-language and the principles specific to UG to very simple forms while maintaining (and in some interesting cases, enhancing) explanatory adequacy, thus approaching what should be true of UG for the reasons indicated above.

This is not the place to try to review these developments, or the range of approaches that have been pursued. My own sense, for what it is worth, is that the basic approach outlined in \textit{Aspects} remains generally viable and can serve as a framework for carrying forward today’s very lively and exciting research endeavors in the study of language and related domains.

\textsuperscript{6} See Susan Curtiss (2012).

\textsuperscript{7} For some recent evidence, see S. Shultz et al. (2014).
References


MIND THE GAP

ROBERT C. BERWICK
Massachusetts Institute of Technology

1 Descriptive and explanatory adequacy in Aspects

Chapter 1 of Chomsky 1965 (henceforth Aspects) doubles down on the goals of linguistic theory.
Not only should we aim for descriptive adequacy – characterizing the possible human grammars – we should try to meet the more stringent demand of explanatory adequacy, reflecting the child’s ability to select a descriptively adequate grammar given the data they receive:

A child who is capable of language learning must have … (v) a method for selecting one of the (presumably infinitely many) hypotheses that are allowed by [the theory of grammar] and are compatible with the primary linguistic data. Correspondingly, a theory of linguistic structure that aims for explanatory adequacy must contain … (v) a way of evaluating alternative proposed grammars… [a] specification of a function \( m \) such that \( m(i) \) is an integer associated with the grammar \( G_i \) as its value (with, let us say, lower value indicated by higher number).

Aspects, pp. 30-31

But what is this “way of evaluating alternative proposed grammars”? By way of illustration Chomsky contrasts two possible “grammars” for a dataset containing the possible sequences of English auxiliary verbal elements. There are eight such sequences: first an obligatory Tense element, then a Modal form or not (will, etc.); followed by a Perfective form (have) or not; then a Progressive (be) or not. Each binary choice allows for 2 possibilities, so there are \( 2^3 = 8 \) auxiliary sequences in all, including the sequence where none of the latter 3 elements is present. These 8 statements, the “data,” can be generated by a grammar G1 with eight rewrite rules, as follows:

(1) a. Aux \rightarrow\text{Tense}
b. Aux \rightarrow\text{Tense Modal}
c. Aux \rightarrow\text{Tense Perfective}
d. Aux \rightarrow\text{Tense Progressive}
e. Aux → Tense Modal Perfective
f. Aux → Tense Modal Progressive
g. Aux → Tense Perfective Progressive
h. Aux → Tense Modal
i. Aux → Tense Modal

However, as Chomsky then describes, if our notational system can use parentheses to denote optionality, all eight data patterns can also be generated a second way, as grammar G2 with just a single rule, (2):

(2) Aux → Tense (Modal) (Perfective) (Progressive)

Since the three items in parentheses can either be present or not, it is straightforward to see that this one-rule grammar captures the same $2^3 = 8$ examples as the eight rule grammar. The two grammars are thus both weakly and strongly equivalent in terms of descriptive adequacy. What about explanatory adequacy? G1 contains roughly the same number of rules that would be required to describe any random sequence of three binary-valued elements. Consequently, this is the worst that a grammar can ever do in terms of explanation, since it fails to capture any regularity in the data. It has simply memorized the data as a list. In contrast, G2 is exponentially more succinct than G1. It is in this sense that G2 has "compressed" the original dataset into a smaller rule system, while G1 has not. This kind of compression may be taken as the hallmark of explanatory adequacy as set out in Aspects. In the remainder of this chapter we show that this kind of exponential reduction in grammar size can indeed serve as a litmus test for discriminating among otherwise descriptively equivalent grammars. Such gaps apparently arise whenever one grammar fails to capture generalizations and then must resort to memorizing – explicitly listing – the data as if it were a random string, while the alternative grammar “compresses” the data set exponentially. It is in this sense that “exponential gaps” indicate that overly large grammars have missed generalizations in the linguistic data. They lack explanatory adequacy.

It is also clear that the ability to compress data depends on the machinery the representational system provides, an empirical matter, again as Chomsky notes. Rule (2) depends on the availability of something like the parenthesis notation to denote optionality. If the auxiliary verb sequences had formed some other set of patterns, for instance, as Aspects notes, the cyclic permutation of \{Modal, Perfective, Progressive\}, then the parentheses notation would not yield a one-rule description of this new dataset. Rather, a system that included different representational machinery, cyclic permutation, would be able to compress these eight new examples down to one rule. In this way, explanatory adequacy all depends on the available representational armamentarium. Consequently, Chomsky rightly emphasizes that what does the heavy lifting in theory-building here is not the particular ‘calculus’ involved – counting symbols in this particular case – but rather the substantive, empirical constraints on the machinery of the mind/brain.

---

1 Actually ternary valued, since there are three possible elements in each position or $3^3 = 27$ possible random strings of length 3 rather than just $2^3$. This corresponds to the number of bits required to encode any string over an alphabet of size 3. While one actually needs bits to more properly account for the coding of the notational system’s alphabet, this detail does not change the comparative advantage of G2 as compared to G1. See Berwick (1982), chapter 4, for further discussion.
The remainder of this chapter is organized as follows. In Section 2 we show that succinctness gaps arise in the case of somewhat artificially constructed examples such as the difference between certain regular and strictly context-free grammars. Section 3 turns to more realistic linguistic territory and establishes that the same kind of exponential size difference holds between the 1970s-1980s models of transformational generative grammar (TGG) and Generalized Phrase Structure Grammar (GPSG): TGGs can be exponentially more succinct than strongly equivalent GPSGs. Intuitively, the reason why is that TGG factors apart movement from underlying phrase structure, while GPSG unnecessarily folds movement into the nonterminal labels of phrase structure rules. More recently, using essentially the same argument, Stabler (2013) has shown that a particular formalization of ‘minimalist’ grammars is exponentially more succinct than an otherwise equivalent extended version of context-free grammars called multiple context-free grammars. Taken together, results like these suggest that one should indeed ‘mind the gap’ between the sizes of alternative linguistic theories that are otherwise descriptively adequate. Section 4 concludes with a comparison of this basic result about the size of grammars to the related systems of evaluation known as minimum description length (MDL) and Bayesian model selection.

2 Regular grammars, context-free grammars, and explanatory adequacy

To see how the “mind the gap” litmus test for explanatory adequacy works in practice, we first consider a simplified, artificial example of two otherwise descriptively adequate theories for the same (finite) language. Section 3 then turns to a more realistic setting. Our artificial example focuses on the finite palindrome language over the alphabet of two symbols, \{a, b\}, with sentences less than or equal some fixed length, \(n\), for even \(n\). For instance, if \(n=4\), then this language encompasses the six sentences, \{aa, bb, abba, baab, aaaa, bbbb\}. We denote this set of sentences by \(L_{\text{PAL}}-n\), with \(n\) even, and let \(|L_{\text{PAL}}-n|\) denote the size of this dataset. Since \(L_{\text{PAL}}-n\) for any even \(n\) is a finite language, it is also regular, and therefore describable by a (deterministic) finite-state automaton (FSA), \(FSA_{\text{PAL}}-n\), as well as by a right-linear, regular grammar (RG), as shown in Figure 1(a). Here the nonterminals in the RG correspond to the states in the FSA, and the total size of the RG is 39 (putting to one side the exact bit length encoding). This language is also equally well describable via a pushdown state automaton (PDA) with 4 states and 10 rules as well as a context-free grammar with 4 rules as shown in Figure 1(b)\(^2\).

\(^2\) The restriction to a finite language is crucial here. If we move to infinite regular languages, then the succinctness gap between finite-state grammars and context-free grammars generating the same language becomes even more enormous – it need not be bounded by any recursive function (Meyer and Fischer, 1971). And in fact, the PDA and grammar in Figure 1(b) will recognize exactly the palindromic sentences of any length. To ensure that sentences longer than \(n\) are not recognized, the CFG should really have to use, e.g., nonterminals \(S_1, S_2\), etc. to ‘count’ the length of the generated sentences. Note however that this only increases the size of the CFG (or PDA) by a linear amount. The grammar or PDA would still be exponentially more succinct than the equivalent RG. To simplify the presentation here, as with the FSA we have not included a special failure state; this would absorb any sentences longer than four or any non-palindromic sequences of length four or less. We have also glossed over details about measuring the RG or PDA/CFG size that are not relevant here. In reality, as noted in the main text, we must count the number of bits encoding the PDA, including its alphabet symbols, transitions, and pushdown stack moves. A more exact accounting in information theoretic terms would also include the probability of each automaton move; see Li and Vitányi (1997) for more details.
While both theories are descriptively adequate, it is not hard to show that as we consider palindrome languages of increasing length, the PDA becomes and exponentially more succinct than the corresponding FSA; similarly for their corresponding grammars. Further, it is easy to see by inspection of the grammar in Figure 1(a) that the finite-state representation in effect has simply listed all the possible strings in $L_{\text{PAL-4}}$, with $|L_{\text{PAL-4}}| \approx |R_{\text{PAL-4}}|$. In contrast, the PDA/CFG does capture the palindrome generalization because the PDA or CFG for this language is smaller than the sentences it describes. The reason why this is so is straightforward. Each time $n$ increases, say from 4 to 6, the FSA or grammar must be able to distinguish among all the possible new ‘left-hand’ parts of each palindrome sentence before the middle, in order to ensure that it can correctly decide whether the following ‘right-hand’ part is the same as the left. For the change from $n=4$ to 6, this means being able to distinguish among all the distinct strings of length 3, the first half of all the new palindromes of length 6, and there are $2^3=8$ of these ($aaa$, $aab$, $abb$, etc). The FSA must be able to do this in order to recognize that the second half of a length-6 sentence is properly paired with the first half. But the only way an FSA can do this is to have distinct states for each one of these $2^3$ possibilities. Therefore, in general, the palindrome-recognizing FSA will have to distinguish among $2^n/2$ strings. Thus as $n$ increases, the number of states in the corresponding FSA or finite-state grammar must increase in proportion to $2^n/2$ – an exponential rate of increase in $n$. Put another way, the FSA must memorize all strings of length $n/2$ or less by using its states.

In contrast, for the palindrome recognizing PDA, increasing $n$ does not demand a corresponding increase in its size, since it can use the same set of fixed general rules that “match up” corresponding pairs of $a$’s and $b$’s and these do not need to be altered at all: the size of the PDA grows linearly with respect to $n$. In this sense, the PDA has captured the essence of the palindrome-type language pattern, matching up paired symbols. In contrast, a similar-sized FSA could have just as well described any set of random strings up to length $n/2$, since the number of states in an FSA is the same for both palindromic strings and any random set of sentences of length $n/2$. In linguistic terms, we say that the FSA or corresponding grammar has “missed the generalization” for the palindrome pattern, because its description – the FSA or grammar size – is almost exactly the same size as the data to be described – that is, proportional to $2^n$.

$$Q_0 \rightarrow aQ_1 \mid Q_0 \rightarrow bQ_2$$
$$Q_1 \rightarrow aQ_3 \mid Q_1 \rightarrow aQ_3 \mid Q_1 \rightarrow bQ_4$$
$$Q_2 \rightarrow aQ_5 \mid Q_2 \rightarrow b \mid Q_2 \rightarrow bQ_6$$
$$Q_3 \rightarrow aQ_7 \mid Q_4 \rightarrow bQ_7$$
$$Q_5 \rightarrow aQ_8 \mid Q_6 \rightarrow bQ_8$$
$$Q_7 \rightarrow a \mid Q_8 \rightarrow b$$

Figure 1(a). A regular grammar and a graphical representation of a deterministic finite-state automaton with 10 states and 12 directed arcs accepting $L_{\text{PAL-4}}$, the palindrome sentences over \{a, b\}, of length up to 4. The automaton starts in state $q_0$. State $q_9$, the double circle, is the final accepting state. The labeled directed arcs between states denote FSA transitions. Note how the
FSA uses distinct states to keep track of all possible left-hand parts of the palindrome sentences, of length up to 2; there are \(2^2=4\) of these.

\[
\begin{align*}
\text{Start} & \rightarrow S \\
S & \rightarrow a \, S \, a \mid \lambda \\
S & \rightarrow b \, S \, b \mid \lambda
\end{align*}
\]

Figure 1(b) A context-free grammar and an equivalent push-down automaton for \(L_{\text{PAL-4}}\) with 4 states and 10 push/pop transition operations that move the PDA between states. The PDA starts in state \(q_0\) and its final state is \(q_3\). The symbol \(\lambda\) denotes the empty string and \(Z\) a special bottom of the stack element. The operations on each arc are in the form input: pop top of stack symbol; new stack contents. For example, the operation \(b, Z; bz\) says that if a \(b\) is the current input symbol, and the result of popping off the top-of-stack is \(a\), then the new stack configuration is \(bZ\), i.e., \(bZ\) is pushed onto the stack, and the PDA moves to state \(q_1\).³

### 3 Linguistic examples: GB, GPSG, and Minimalist grammars

The RG vs. CFG example of the previous section may appear artificial, but it does bear some resemblance to the pairing of Subject NPs and verbs in ‘nested’ or center-embedded sentences, demonstrating that two descriptively adequate theories can have radically different sizes, and illustrating that descriptive succinctness serves as a good litmus test for successfully capturing generalizations in data. More importantly however, one can also demonstrate this same effect in a more realistic setting, in the analysis of Generalized Phrase Structure Grammar (GPSG), as first observed in Berwick (1982).⁴ Similar to the way in which FSGs can record finite palindrome patterns, but at a cost of an exponential growth in size of the resulting automaton/grammar when compared to PDAs/CFGs, GPSGs can record filler-gap dependencies via an expanded set of nonterminals, with a comparable exponential growth. As before, this indicates that the GPSG description has missed some generalization about filler-gap dependencies, and has simply listed the possibilities where a more compact representation is possible.

---

³ To simplify the presentation, as in the FSA we have not included a special failure state; this would adsorb any sentences longer than four or any non-palindromic sequences of length four or less. The PDA as presented will of course accept all and only the palindromes of any length.

⁴ An earlier version of this material first appeared in Berwick (1982), chapter 4. Details of the results described in this section may be found in this chapter. The main result that GPSGs grow exponentially in size with the number of ‘filler’ items that may be displaced, is established as theorem 4.1 there.
To demonstrate the exponential blow-up associated with GPSG and filler-gap dependencies, we first recall how GPSG encodes a single filler-gap dependency through an expanded set of nonterminals. We begin with a sentence that has all its ‘fillers’ in canonical argument positions, e.g.,

(3) John kissed Mary

As usual, an object *wh*-question version of this sentence has a ‘filler,’ e.g., *who*, at the front of the sentence, linked to the ‘gap’ position as the argument of *kissed*:

(4) Who [did John kiss gap ]

The GPSG analysis of this sentence links the filler *who* to the (unpronounced) gap via an analysis where ordinary nonterminals like S, VP, or NP have ‘slashed’ counterparts S/NP, VP/NP, and NP/NP, where XP/YP is to be interpreted as a constituent tree of type X that dominates (and is missing) a constituent of type YP expanded as an unpronounced element somewhere below (Gazdar, 1981). ‘Slash’ rules are introduced by context-free rules of the usual sort, but using the ‘slashed’ nonterminal names, e.g.:

(5) Comp → Who S/NP

In this way, one can arrange for a sequence of ‘slash’ rules to serve as a way to ‘store’ the information that a *wh*-NP is at the front of the sentence, as in (5), and pass that information down a (c-commanding) chain of nonterminal expansions, terminating with the rule NP/NP → ε, where ε is the empty string (the gap). In the terminology of contemporary transformational grammar, the GPSG rules construct something like a chain, with its head positioned at the filler, and then the encoding of that filler as intermediate slashed nonterminals, ending in a rule that expands as the empty string in the ‘gap’ position. The resulting syntactic tree would look something like the following in Figure (3) where irrelevant syntactic details have been omitted:

```
CP
   NP+wh
      Who
         did John
             VP/NP
               V
                 NP/NP
                   kiss
                        ε
```

Figure 3: Slashed nonterminals for the sentence *Who did John kiss* link a *wh*-NP filler to its ‘gap’ position.

The question of succinctness of this representation comes into play when there are multiple ‘displacements’ from lower clauses to higher ones, in examples like the following, where a *wh*-NP filler *which violins*, fills a gap position after *on*, while a second NP filler *these sonatas* is the argument of *to play*:
(6) [Which violins], are [these sonatas], difficult to play [ ] on [ ]? How are we to deal with this kind of example, with two active ‘fillers’ via the slash-nonterminal representation? We can extend the ‘slash’ notation to explicitly carry these along via the nonterminal names, e.g.:

\begin{align*}
&7) \ CP \rightarrow \text{wh-NP}_1 (\text{which violins}) \ S/\text{NP}_1; \ S/\text{NP}_1 \rightarrow \text{NP}_2 (\text{these sonatas}) \ S/\text{NP}_1\text{NP}_2; \\
&\quad S/\text{NP}_1\text{NP}_2 \rightarrow (\text{pro}) \ \text{VP/\text{NP}_1\text{NP}_2}; \ \text{VP/\text{NP}_1\text{NP}_2} \rightarrow \text{V/\text{NP}_2/\text{NP}_2} \ \text{PP/\text{NP}_1}; \ \text{NP}_2/\text{NP}_2 \rightarrow \varepsilon; \\
&\quad \text{PP/\text{NP}_1} \rightarrow \text{NP}_1/\text{NP}_1; \ \text{NP}_1/\text{NP}_1 \rightarrow \varepsilon;
\end{align*}

Consequently, in addition to ‘single’ slashed categories such as S/\text{NP}_1, handling multiple fillers requires additional categories like S/\text{NP}_1\text{NP}_2 (along with subsequent chains VP/\text{NP}_1\text{NP}_2 etc.), in effect listing the fillers and their order (here, via indices) so that they can be ‘discharged’ in the ‘gap’ positions in which they are to be interpreted. But in general, there might be any number of such ‘displaced’ constituents $n$ as the length of sentences grows, and one could potentially choose to ‘move’ a filler from its canonical argument position or not, as well as possibly overlapping filler and gap chains. That is, it might be that the order S/\text{NP}_2\text{NP}_1 is sometimes be more acceptable than S/\text{NP}_1\text{NP}_2 (Stowell, 1982).\footnote{Fodor (1978) argues that the filler-gap order must always follow a nested, pushdown stack form, but this result is open to question; see Berwick (1982), Chapter 3.} Without any additional constraints then, having to handle $n$ displaced fillers in any possible order (including no displacement), implies that in the worst case a descriptively adequate GPSG for such languages would require distinct slashed nonterminals and rules for any of the possible $2^n$ subsets of the $n$ fillers, an exponential number. Thus the size of GPSGs to handle filler-gap relations grows exponentially with the number of filler-gap pairs, the usual warning sign of a failure to capture some generalization, as with the FSG case described in Section 2.\footnote{Theorem 4.1 in Berwick (1982) has a more detailed account of this result. The approach there uses (deterministic) frontier-to-root tree automata, and shows that if there were a non-exponential size GPSG for handling filler gap relations via new nonterminals, then this would imply that there is a non-exponentially sized FSA for recognizing the palindrome languages of a given length, shown to be impossible in the main text. This 1982 result should be compared to the one in Stabler (2013) described immediately below.}

Put another way, encoding derivational ‘movement’ via the introduction of new nonterminal names is possible, but expands the grammar size unnecessarily: every intervening nonterminal and context-free rule along the nonterminal path from filler to gap is affected and requires slash-category additions, when in fact these intervening constituents almost always do not play any role in the filler-gap relation. The expansion in terms of rules is really unnecessary. In contrast, a GB-based account introduces no such unwarranted expansion in the number of nonterminals and rules, because it does not ‘record’ at each intermediate nonterminal the possible combinations of displaced constituents. These intervening elements are irrelevant to the separate rule of ‘Move alpha,’ which does not make reference to these intervening nonterminals (apart from a fixed set of constraints that do play a role, such as whether an intervening nonterminal is a bounding node like CP, landing sites, and the like). Another way to state this result is that GPSG has ‘multiplied out’ all the possible phrase structure chains that can intervene between fillers and gaps, even though the identity of the particular phrase structure elements involved does not really matter. In contrast, GB keeps these two components separate, at a real savings in grammar size.
More recently, an exponential succinctness gap has also been demonstrated by Stabler (2013), when comparing the relative grammar sizes of a formalization of ‘minimalist’ grammars (MGs) to an extension of context free grammars known as multiple context-free grammars (MCFGs). Stabler demonstrates that two otherwise strongly equivalent MGs and MCFGs can differ exponentially in size: the MCFG contains exponentially more rules than the equivalent MG – and for the same reason as with the GPSG/GB contrast in Section 2. In an MG, the role of ‘movement’ is taken over by an operation called ‘internal merge’ – that is, the operation of merge between two syntactic (set-formulated) objects where one, the constituent that would be the target of move alpha in GB theory and the “moved constituent,” is a subset of the other. The output is a new syntactic object where the moved constituent is copied to its new ‘filler’ position – there is no ‘gap left behind which would be occupied by a phonologically empty element in older GB theory, but simply the original constituent. 8(a-b) illustrate the syntactic structure before and after an internal merge operation (irrelevant details suppressed); note that the second argument to the merge operator, [DP what], is contained in (is a subset of) the first argument.

(8a) Merge( [CP people [VP eat [DP what]]], [DP what] ) → 
(b) [CP [DP what] [people [VP eat [DP what]]]]

Aside from the initial position of the constituent what and its ‘landing site’, this movement (actually copying) does not need to refer to any intermediate syntax in between (such as VP). If there is more than one ‘mover’, then this simply adds a single new lexical entry, as described in more detail in Stabler (2013). In contrast, in MCFGs, such effects are handled by the use of variables associated with nonterminals that can hold strings such as what. For example, the nonterminal VP(x₁, x₂) stands for an ordinary VP that has two variables that can hold the values of two strings, x₁ and x₂. The first variable can correspond to the verb eat, while the second can correspond to what. The following MCFG rule indicates that the value of the first variable – the verb – can be associated with the V nonterminal, while the second variable holding what can be associated with the DP:

(9) VP(x₁, x₂) → V(x₁) DP(x₂)

This wh element variable can then be ‘passed up’ through each higher nonterminal until what reaches its landing site in the CP, as per the following MCFG rule. Note in particular that in the MCFG nonterminal CP that the order of the two variables is reversed and their values concatenated together into one string x₂ x₁, which places what before the verb:

(10) CP(x₂ x₁) → C VP(x₁, x₂)

It should be apparent that the MCFG proceeds just as with GPSG when it encodes movement, in that all intervening nonterminals are modified along the chain from the initial position of what all the way to its landing site – as before, affecting intervening nonterminals by introducing variables even though those nonterminals are not really implicated in the ‘movement.’ As a result, if there are multiple possible ‘movers,’ each corresponding to a different kind of lexical item or feature, then the MCFG must multiply out all these possibilities as with GPSG. This amounts to an exponential number of choices since the possible movements can be any of the subsets of the movers, and each might move or not, as Stabler notes. Stabler concludes: “MGs can be exponentially smaller than their strongly equivalent MCFGs because MCFGs explicitly
code each movement possibility into the category system, while MGs can, in effect, quantify over all categories with a given feature.” This exponential succinctness gap between MGs and MCFGs again serves to signal that MGs can capture a generalization that the MCFGs cannot, in this case, a lexical generalization that is quantified over all nonterminals with a particular feature.

4 Conclusion: explanatory adequacy and aspects of simplicity

Putting together the results from the previous sections, we see that a grammar’s relative succinctness can be one way to fruitfully probe into whether a linguistic theory is able to capture natural generalizations or not, exactly as anticipated in Aspects. We conclude here by showing that this approach is in line with two other approaches to “evaluating” grammatical theories, one called “Minimum Description Length” (MDL, Rissanen, 1978), and the other, Bayesian model selection.

We first consider MDL. Suppose we have a family of grammars \( \mathcal{G} \) and a given set of sentences, a ‘corpus,’ \( D \). MDL defines the ‘best’ grammar \( G \) over some family of grammars \( \mathcal{G} \) as that grammar which minimizes the sum of two components. The first is \( |G| \), the size of the grammar as measured in bits. This factor is perhaps the one most familiar to linguists. The second factor is \( |D_G| \), the size of the data as encoded or generated by \( G \). The intuition behind this measure is that a good theory will be able to “compress” the original data \( D \) such that \( |D_G| \) is smaller, usually much smaller, than \( D \) itself. If \( D \) is not compressible, then there is no smaller description of the data by any theory aside from a listing of the data itself. Note that if there are some data examples that the given grammar cannot generate – exceptions to rules, in traditional terminology – then these must be added in to the \( |D_G| \) size factor by an explicit listing of these examples without any compression. We can write out the MDL formulation in the following way:

\[
(10) \quad |G| + |D_G|
\]

Conventionally, this is given as the following optimization problem:

\[
(11) \quad \arg\min_{G \in \mathcal{G}} \ |G| + |D_G|
\]

Without working through all the details, this MDL approach will yield the same “exponential gap” litmus test as described in Sections 2 and 3: exponential gaps in grammar size will show up in the first \( |G| \) factor, while explicit listing of data – if one is not using an appropriate notational framework – will show up in exponential gaps in the second factor. For one way of using this MDL framework in a concrete linguistic application, the inference of morphological and syntactic regularities starting from strings of phonemes and proceeding through to syntax, see de Marcken (1996). More recently, MDL has been explicitly incorporated in several other models of language acquisition, e.g., using categorial grammars (Villavicencio, 2002); or slightly augmented context-free grammars (Hsu and Chater, 2010). A related approach to MDL uses the

---

7 One might rightly enquire as to whether the number of sentences in \( D \) is infinite or not (see Note 2). For our purposes here this does not matter. For one way of systematically handling this question, using the notion of “uniform computability” as in the case of a sequence of fixed Boolean circuits, see Berwick (1982).
The notion of program size complexity: it attempts to find the length of the shortest program that can compute or generate a particular language. Chater et al. (2003) and Hsu, et al. (2013) discuss this approach in the context of language learnability; more on its relationship to linguistic theories can also be found in Berwick (1982).

Further, the MDL approach itself can be closely identified with Bayesian evaluation methods as first pioneered by Horning (1969), who used Bayesian inference to select grammars within an acquisition-by-enumeration approach. In a Bayesian framework, we fix some set of (observed) data $D$ and some description of $D$ via a class of (stochastic) context-free grammars, $\mathcal{G}$, that can generate $D$, with some prior probability. A Bayesian inference approach would then attempt to find the particular grammar $G \in \mathcal{G}$ that maximizes the posterior probability of $G$ given the data $D$, i.e., $\arg\max_{G \in \mathcal{G}} p(G | D)$. Letting $p(G)$ denote the prior probability assigned to the (stochastic) context-free grammar $g$, we can use Bayes’ rule in the usual way to find this maximum by computing $p(G|D) = \arg\max_{G \in \mathcal{G}} p(D|G) \times p(G)/p(D)$. Since $D$ is fixed, it can be ignored to find the $G$ that maximizes the product, and we have the usual computation that attempts to find the grammar $G$ that maximizes the product of the prior probability of $G$ times the likelihood of the data given some particular grammar, $p(D|G)$. In other words, Bayesian inference attempts to find the $G$ that satisfies the following formula:

$$ (12) \quad G = \arg\max_{G \in \mathcal{G}} p(D|G) \times p(G) $$

To pass from this formulation to the MDL version one way to proceed is as follows. The MDL principle as applied to stochastic context-free grammars says that the ‘best’ grammar $G$ minimizes the sum of the description length of the grammar and the description length of the data given $G$. More precisely, if $|G|$ is the length of the shortest encoding of grammar $G$ and $|D_G|$ is the length of the shortest encoding of the data $D$ given the grammar $G$, then MDL attempts to find:

$$ (13) \quad G = \arg\min_{G \in \mathcal{G}} |G| + |D_G| $$

Using near-optimal coding schemes, Shannon’s source coding theorem (1948) implies that the description of the length of $D$ with respect to a particular grammar $G$ can be made to closely approach the value $-\log_2 p(D|G)$. We can further assume, as is standard, that one way to define the prior probability of a stochastic context-free grammar $p(G)$ is as $2^{-|G|}$. Larger grammars are penalized in the same sense that we have used throughout this chapter and have lower probability. Taking $\log_2$ of (8) we get:

$$ (14) \quad G = \arg\max_{G \in \mathcal{G}} \log_2 p(D|G) + \log_2 p(G) $$

Substituting $-|D_G|$ for $\log_2 p(D|G)$ and $2^{-|G|}$ for $p(G)$ and pulling out the negative sign we get:

---

8The notion of program size complexity was developed independently by Solomonoff (1960, 1964) and Kolmogorov (1965); for a recent comprehensive survey of this field, see Li and Vitányi (1997). In this framework one can show that there is an ‘optimal’ universal programming system (a representation language or grammar) in the sense that it is within a constant factor of any other optimal programming language. A detailed analysis of this quite useful approach lies beyond the scope of this chapter; for a concrete example in the context of language acquisition, see Hsu and Chater (2010).
\[
(15) \ G = \ \arg\max_{G \in \mathcal{G}} -|D_G| + (-|G|) = \ \arg\min_{G \in \mathcal{G}} |D_G| + |G|
\]

In this particular case then, the MDL formulation (13) and the Bayesian formulation (14) coincide.\textsuperscript{9}

It should be emphasized that this is not the only notion of “explanatory adequacy” that might prove valuable in choosing among otherwise descriptively adequate linguistic theories. Other approaches might stress the computational complexity of acquisition – the sample complexity or the number of examples required to acquire language (Berwick, 1982, 1985); or the computational complexity associated with the use of language, that is, parsing or production.

Summarizing, we have found that three different ways to formulate grammar evaluation in light of explanatory adequacy all amount to the same kind of calculation, ultimately grounded on the notion of the size of a grammar plus the size of the linguistic data as encoded by that grammar. Further, this measure can be applied to actual alternative theoretical proposals in linguistics, distinguishing between proposals that offer generalizations of data as opposed to those that do not. Finally, this analysis shows that currently popular approaches to learning in cognitive science, such as Bayesian methods, turn out to be worked-out versions of explanatory adequacy as discussed in Aspects. In this respect, contrary to what is sometimes thought, the informal notion of size and simplicity as litmus tests for linguistic theories can be placed within a coherent framework. We take all this as lending support to the view, first stated in Aspects, that explanatory adequacy has an important role in evaluating linguistic theories, incorporating a view that one must attend to how grammars are acquired or inferred from data.

References


\textsuperscript{9} This Bayesian estimation is usually called a ‘maximum a posteriori estimate’ or MAP estimate. Note that a Bayesian account would give us much more than this, not only just the maximum a posteriori estimate, but also the entire posterior distribution – the whole point of Bayesian analysis, on some accounts.


Questions with Declarative Syntax Tell Us What About Selection?\(^1\)

Jonathan David Bobaljik  
Susi Wurmbrand  
University of Connecticut

1 Introduction

One of the many enduring themes of Chomsky's (1965) *Aspects* is the question of selection (broadly construed) and the distinction among syntactic and semantic properties (features) of linguistic expressions. In this brief contribution, we aim to reaffirm the role that syntactic selection plays in the domain of clausal embedding; that is, where verbs select for a complement of a particular syntactic type and a semantically (or pragmatically) equivalent utterance is sharply ungrammatical. Our specific focus is to synthesize a body of literature on the phenomenon of ‘optional’ (non-echo) *wh-in-situ* in *wh*-movement languages, arguing ultimately that syntactically, the phenomenon as such may not exist. What appears to be *wh*-in-*situ* in these languages may carry interrogative force as a speech act, but from a syntactic perspective is a declarative clause with a *wh*-expression in focus—a question with declarative syntax (DSQ). The key evidence for this claim comes from selection/subcategorization. The relevant facts have been noted for individual languages, including English, but we offer here a meta-study, of sorts, contending that the generalization in (1) holds systematically across all languages we have been able to examine, despite a wealth of variation along other dimensions.

(1) DSQ/*wh-in-situ* generalization:  
If a language has *wh*-movement (to Spec,CP), then *wh*-movement is obligatory in indirect questions. Equivalently: If a *wh*-movement language allows ‘optional’ *wh-in-situ*, the *in-situ* construction is blocked in selected questions.

\(^1\) For examples from other languages we thank Meltem Kelepir (Turkish), Gíslí Rúnar Harðarson (Icelandic), and the ARD and ZDF Mediatheken. The paper also greatly profited from feedback and discussions with Benjamin Girard-Bond, Željko Bošković, Jon Gajewski, Magdalena Kaufmann, Jairo Nunes, Lara Reglero, Koji Shimamura, Sandra Wood, and especially Cynthia Levart-Zocca.
The systematicity with which this holds, and the equally systematic absence of a restriction on embedding in *wh-in-situ* languages, provides a compelling argument, in our view, against analyses such as Cheng and Rooryck (2000) and Pires and Taylor (2007) which propose to assimilate the optional *wh-in-situ* facts of English, French or other languages to the *wh-in-situ* constructions of, say, Chinese, Japanese or Turkish. We suggest instead (partly in line with Ginzburg and Sag’s [G&S] 2000 analysis of English) that DSQs are syntactically, if not semantically, akin to echo questions. In contrast to true *wh-in-situ*, DSQs involve no interrogative syntax (no question operator or interrogative C) and find an interpretation as a question in the pragmatics (for accounts of echo questions along these lines, see Artstein 2002, Poschmann 2010). True *wh-in-situ* by contrast (contra G&S) involves an interrogative complementizer, C_{WH}, and thus is syntactically typed as a question (Cheng 1991; whether this triggers covert movement in one form or another, or merely binds the *wh*-word is immaterial here). This C_{WH} allows (true) *wh-in-situ* questions to be selected by a higher predicate, forming indirect questions. Since DSQs lack C_{WH}, not only is there no *wh*-movement (overt or covert), but a DSQ clause cannot be selected as an interrogative, explaining the observed correlation. At its core, our explanation of (1) is thus that selectional compatibility reveals syntactic properties that are partially obscured in simple clauses by alternative (“pragmatic”) strategies. In their syntax languages are either *wh*-movement or *wh-in-situ* (more accurately, *wh-in-FOCUS*) but no language (that we know of) truly mixes both constructions. As a syntactic corollary, we note that the generalization presented here supports an account where *wh*-movement, when it occurs, is the consequence of a property of C, and not triggered by a property inherent to *wh*-expressions forcing them to move (or to be syntactically licensed in any way). We sketch a formal account within the Reverse Agree framework of Wurmbrand (2012a, b, 2014) which supports the various assumptions we are led to from the broader, cross-linguistic considerations just sketched.

2 DSQs

English is routinely described as lacking *wh-in-situ*, and thus requiring movement of a *wh*-expression in contexts other than echo questions. This is an over-simplification. In English, it is possible to ask an information-seeking, non-echo *wh*-question without fronting the *wh*-word, as noted, for example, in standard descriptions (Huddleston and Pullum 2002: 973; hereafter *CGEL*) and occasionally in the theoretical literature (Postal 1972, G&S 2000, Pires and Taylor 2007, Zocca DeRoma 2011). Some examples are given in (2); none of these fit the defining criteria of echo questions in the contexts in which they occur.²

2 Examples (2c,e) are from the collection of examples in Zocca DeRoma (2011).

(2)  a. So, your boy’s name is **what**?  [McNulty, *The Wire*, season 1, episode 1]
    b. Major, you want this stuff **where**?  [unnamed officer, *The Wire*, season 2, episode 1]
    c. A: All the creative people — our R&D, marketing, in-house ad staff — that's all done here in Jersey.
        B: But the sneaks are made **where**? China? Malaysia?  [Stabler, *Law and Order*]
    d. Briscoe: What do you suspect?
        Doctor: She was poisoned, by her own hand or someone else’s.
        Briscoe: And you know this **how**?
Questions with Declarative Syntax

Doctor: Fresh needle mark on her left buttock. [Law & Order, season 10, episode 23]

e. Mrs. Valentine: Good news, darling. Your father's spirit guide has allowed him to
leave his crystals and meditation mat long enough to come to New York on business,
and he wants to see you tonight.

Drue: And this is good news on what planet? [Dawson’s Creek, Season 4, Ep. 11]
f. “Now,” said Umbridge, looking up at Trelawney, “you’ve been in this post how long,
exactly?” [Harry Potter and the Order of the Phoenix, Ch. 15]

Some German examples illustrating the same point are given in (3):

(3) a. Und diese Teilhaber erreichen wir wie?
‘And we can reach these partners how?’ [Stark, Schweinegeld / Tatort ep. 746]

b. Köhler: Ich hab nicht viel Geld. Herr Klarmann hat mir geholfen…
Wilsberg: Und diese “Hilfe” dauert jetzt wieviel Jahre?
‘And this “help” has been lasting for how many years?’ [Wilsberg, Bullenball]

c. Du bist aus welchem Grund nach Patmos gefahren?
you are for which reason to Patmos gone
‘You went to Patmos for what reason?’ [Bayer 2006]

Though somewhat restricted to a colloquial register, examples of DSQs are not hard to find. They are distinct from echo questions, in that, for example, they occur in contexts that are not asking for repetition or clarification of an expression in a preceding utterance (see CGEL: 886-891). The examples in (2a) and especially (2b,d), as is clear from the contexts they occur in, are genuine requests for information. The early literature identifies, as species of unmoved questions in English, at least legalistic questions (Postal 1972) which further a string of questioning such as (2a,c), and quiz-show questions, such as (4). (4)

a. And now, for $5,000, London is the capital of which country? (Cooper 1983:148)

b. During the Korean War, the United Nations forces made largely of the troops of the
United Nations and South Korea fought against the troops [of] North Korea and what
[country]?

[c. Louis XIV was how old when he became King? studystack.com

Another context supporting such questions in English, noted by G&S 2000, is one in which further information is requested to flesh out a salient context, where there is no strict linguistic antecedent to count as an echo. Pires and Taylor (2007) illustrate with an example like (5a), while Poschmann (2010) and Kaufmann and Poschmann (2011) give (German) examples like (5b):

(5) a. [Seeing somebody reading]: You’re reading what?

3 Echo questions repeat a prior utterance, substituting a wh-expression such as what for a part of that utterance, requesting repetition or clarification (or rhetorically expressing surprise). In response to A’s utterance in (i), speaker B’s responses in either (ii) or (iii) are echo questions. Note that the string replaced by what in an echo question may be smaller than a word (Cooper 1983:150, CGEL)—this is not generally possible in the (other) DSQs we consider.
b. [Discussing pot-luck plans]: Diane’s baking a cake, Magda’s buying bagels, and Harry’s bringing what?

DSQs in English often have a sarcastic or disdainful edge to them and can be used as rhetorical questions. Examples (2c-f) illustrate this flavor, but this is neither obligatory nor unique to \textit{wh-in-situ} and would be available to the corresponding fronted questions as well.\footnote{The text after (2f) draws this out: “Professor Trelawney scowled at her, arms crossed and shoulders hunched as though wishing to protect herself as much as possible from the indignity of the inspection. After a slight pause in which she seemed to decide that the question was not so offensive that she could reasonably ignore it, she said in a deeply resentful tone, ‘Nearly sixteen years.’”}

In sum, despite their declarative (focus) syntax, DSQs are (or can be) genuine interrogative speech acts. In terms of their syntactic distribution, they are quite free. The apparently \textit{in-situ} \textit{wh}-word may be in an embedded clause, as in (6a), and even in an island (where the corresponding \textit{wh}-movement would be impossible), as in the coordinate structure island in (6b) or the relative clause island in German (6c):

(6) a. And the defendant claimed that he was standing \textit{where}?

b. During the Korean War, the UN forces made largely of the troops of the UN and South Korea fought against the troops of [North Korea and \textit{what country}]?

c. Er hat den Mann, der \textit{aus welchem Grund} nach Patmos gefahren ist, angerufen?

He has the man who out which reason to Patmos gone is called ‘He called the man [who went to Patmos for what reason]?’

A prominent line of analysis (we discuss others below) treats DSQ in English, German, and other \textit{wh}-movement languages as instances of the \textit{wh-in-situ} constructions familiar from languages such as Chinese and Turkish:

(7) a. Hufei mai-le \textit{shenme}? \\
Hufei buy-ASP \textit{what} \hspace{1cm} \textit{Mandarin} \hspace{1cm} \textit{Cheng 2003: 103}

‘What did Hufei buy?’

b. Ozan \textit{ne} oku-du? \\
Ozan what read-PAST \hspace{1cm} \textit{Turkish} \hspace{1cm} \textit{M. Kelepir, p.c.}

‘What did Ozan read?’

For example, Cheng and Rooryck (2000), analyze French ‘optional \textit{wh-in-situ}’ as involving a null interrogative complementizer that licenses the \textit{in-situ} \textit{wh}-expressions in French in a manner directly analogous to \textit{wh-in-situ} licensing in Chinese. Pires and Taylor (2007) offer such a proposal for English (and analogous examples in Brazilian Portuguese). Noting the island-insensitivity, they adapt a familiar analysis of \textit{wh-in-situ}, positing a null C which binds the \textit{wh}-words in its domain, requiring no movement (neither overt nor covert). In these analyses, languages that allow ‘optional’ \textit{wh-in-situ} have two interrogative complementizers, one that is associated with (overt) \textit{wh}-movement, the other licenses the \textit{wh}-word without overt movement, (either with covert movement of one sort or another, or via unselective binding or equivalently, Agree). G&S (2000) propose a very different syntactic treatment of English DSQs, yet share with the analyses above the general suggestion that DSQs in a language like English invoke the same grammatical pieces as standard interrogatives in a \textit{wh-in-situ} language. In the next section,
we show that despite their syntactic freedom, DSQs obey one strikingly robust restriction—they are infelicitous as indirect questions, which sets them apart from true wh-in-situ.

3 Wh-in-situ vs. wh-in-situ

A systematic point of difference between ‘optional’ and ‘true’ wh-in-situ, as far as we have been able to determine, lies in the possibility of occurring as the complement to a predicate that syntactically selects for an interrogative: DSQs cannot occur as indirect questions. In optional wh-in-situ languages, if a wh-phrase occurs in the complement of an interrogative-selecting predicate, wh-movement is obligatory in the embedded clause and unmoved variants are sharply ungrammatical or parsed as direct quotes. This is shown in (8) for English (see also G&S 2000; CGEL: 973), and (9) for German.5

(8) a. *He asked me your boy’s name is what.
   b. *I wonder I should put this stuff where.
   c. *Umbridge asked Trelawney she’s been in the post how long.

(9) a. *Stark hat gefragt diese Teilhaber erreichen wir wie?
   Stark has asked these partners reach we how
   ‘Stark asked we can reach these partners how?’
   b. Stark hat gefragt wie wir diese Teilhaber erreichen.
   Stark has asked how we these partners reach
   ‘Stark asked how we can reach these partners.’

In true wh-in-situ languages, on the other hand, this restriction is not found, and indirect questions show the wh-in-situ configuration in the embedded clause:

(10) a. Botong xiang-zhidao [ Hufei mai-le shenme ]? Mandarin
   Botong want-know [ Hufei buy-ASP what ]
   ‘Botong wants to know what Hufei bought.’
   [Cheng 2003: 103]
   b. [ Ozan’in ne oku-dug-un-u ] merak ed-iyor-um Turkish
   [ Ozan what read-NOM-ACC ] wonder do-imperf-1sg
   ‘I wonder what Ozan read.’
   [M. Kelepir, p.c.]
   c. [Masao-ga [CP Hanako-ga dono hon-o tosyukan-kara Japanese
   [Masao-NOM [CP Hanako-NOM which book-ACC library-from
   karidasita ka ] siritagatteiru ] koto
cHECKED-OUT-Q ] want-to-know ] fact
   ‘the fact that Masao wants to know which book Hanako checked out from the library’
   [Saito 1992: 84, (33a)]

Facts such as these demonstrate that there is nothing in principle preventing a wh-in-situ clause from occurring as the complement to an interrogative-selecting predicate. If (true) wh-in-situ in-

5 Since verb-second distinguishes main from embedded clauses, the direct quote parse of these examples, marginally available in English, is not an interfering factor in German (9a).
volves a null C_{WH}, then that C_{WH} (or the projection it heads) satisfies the selectional requirements of an embedding predicate. The question, then, is why the lack of movement in English and German correlates with an inability to host an indirect question construal. To the extent this question has been addressed, the general proposal (with the notable exception of Bošković 2000 on French, see below) is that the correlation is spurious – an accidental coincidence in English (G&S 2000) and French (Cheng and Rooryck 2000). A cross-linguistic survey (summarized below) shows instead that the correlation appears to be systematic, and thus in need of a principled explanation. A number of languages are syntactically like English and German, in that DSQs look, in their gross syntax, like declarative-selecting sentences. In all these languages, DSQs cannot serve as the complements to interrogative-selecting predicates.

Perfunctory inquiries suggest Dutch and Icelandic are like English and German in the relevant respects. For space reasons, we only include Icelandic examples here.\(^6\)

(11) a. (a classroom situation)
...og Jón Sigurðsson fæddist hvæð?
...and Jón Sigurðsson be.born when
‘...and Jón Sigurðsson was born when?’
b. (a police questioning situation)
... og þú varst hvær þegar Jóna keyrði útaf?
...and you were where when Jóna drove out.of
‘...and you were where when Jóna drove off the road?’

As in English, DSQs are possible in islands, such as coordination (12a) where overt wh-movement is robustly impossible (12b).

(12) a. Sigur Rós hélt 14 tónleika í Bandaríkjunum og hvaða landi?
Sigur Rós held 14 concerts in USA and what country?
‘Sigur Rós held 14 concerts in the US and what country?’
b. *Hvaða landi hélt Sigur Rós tónleika í Bandaríkjunum og hvaða landi?
what country held Sigur Rós concerts in USA and what country
*‘What country did Sigur Rós hold concerts in the US and what country?’

And as in English, DSQs are possible as matrix questions, even in embedded clauses, as in (13a), but in indirect questions, fronting is obligatory (cf. (13b-c)).

(13) a. Obama lét í ljós að hann hefði verið fæddur hvær?
Obama let in light that he had been born where
‘Obama revealed that he was born where?’
b. *Þingmaðurinn spurði Obama væri fæddur hvær?
senator.the asked Obama was born where
‘The senator asked Obama was born where?’
c. Þingmaðurinn spurði hvær Obama væri fæddur?
senator.the asked where Obama was born
‘The senator asked where Obama was born?’

\(^6\) Our thanks to Gísli Rúnar Harðarson for these examples. DSQs are perhaps not as freely available in Icelandic as they are in English, but the relevant contrast is evidently sharp, just as it is in English.
Wood (2009) reports the same effects for American Sign Language (ASL). ASL is especially interesting since there is both leftwards and rightwards movement of wh-expressions with some debate as to the analysis (see Petronio and Lillo-Martin 1997, Neidle et al. 2000). Nevertheless, ASL is like English in allowing apparent optional wh-in-situ as non-echo interrogatives (14b), including in embedded clauses (15b), but this option is prohibited in selected questions (16b), where movement in the embedded clause is obligatory.

(14) a. **WHO** JOHN SEE YESTERDAY?
b. JOHN SEE **WHO** YESTERDAY?
   ‘Who did John see yesterday?’

(15) a. **WHO** JOHN FEEL MARY KISS?
b. JOHN FEEL MARY KISS **WHO**?
   ‘Who does John think (that) Mary kissed?’

(16) a. JOHN ASK (MARY) **WHO** SEE?
b. *JOHN ASK (MARY) SEE **WHO**?
   ‘John asked Mary who she saw.’

Beyond Germanic and ASL, the Romance languages are well known for having apparently optional wh-in-situ to one degree or another, and there is a significant literature on this topic. Despite a wealth of interesting variation to which we cannot do justice, the core main/embedded asymmetry in (1) appears to hold wherever it is testable. Pires and Taylor (2007) and Zocca DeRoma (2011) discuss (Brazilian) Portuguese in comparison to English. They note that English is regularly described as having ‘obligatory’ wh-movement outside of echo contexts, while Portuguese is described as having optional wh-in-situ, as in (17):

(17) a. **Quem** você viu?
    who you saw
b. Você viu **quem**?
    you saw who
   ‘Who did you see?’

Yet closer inspection shows that the languages are essentially the same in their syntax, differing primarily in that the range of pragmatic contexts that permit DSQs. Important for our concerns is the observation that both languages disallow DSQs as indirect questions (18).

(18) a. O Pedro perguntou **quem** você viu?
    The Pedro asked who you saw
   ‘Pedro asked who you saw.’
b. *O Pedro perguntou você viu **quem**?
    The Pedro asked you saw who
   ‘Pedro asked who you saw.’

---

7 The one syntactic difference is that BP allows apparent matrix questions with an overt complementizer. DSQs are excluded in such cases. An explanation may be that these constructions involved a concealed embedding predicate.
Like English, there is no general prohibition on *wh*-in-*situ* in embedded clauses (19a), or even in islands—(19b) permits a reading of the adjunct *wh*-word as a modifier of the predicate in the adjunct clause (how did she fix it?), a reading that is impossible for an overtly moved *wh*-word (19c).

(19) a. *O João pensa que a Maria viu quem?*  
The João thinks that the Maria saw who  
‘John thinks Maria saw who?’

b. *O Pedro saiu depois que a Maria consertou o carro como?*  
The Pedro left after that the Maria fixed the car how  
‘Pedro left after Maria fixed the car how?’ (ambiguous)

c. *Como o Pedro saiu depois que a Maria consertou o carro?*  
How The Pedro left after that the Maria fixed the car  
‘How did Pedro leave after Maria fixed the car?’ (matrix ‘how’ only)

Pragmatically, Brazilian Portuguese is freer in the range of contexts in which DSQs are permitted. Zoeca DeRoma collected examples from Brazilian TV shows (and other sources) including their contexts of occurrence, and it is clear that examples in BP are felicitous in contexts where they are not in colloquial English. Despite this pragmatic variation, there is no appreciable syntactic variation in the key properties of DSQs.

Given the overwhelming prohibition of *wh*-in-*situ* in embedded interrogative clauses, the DSQ/*wh*-in-*situ* generalization in (1) thus appears to be correct and we offer an account of this generalization in the next section.

### 4 DSQ: An account

#### 4.1 The system

In this section, we provide a syntactic account within an explicit feature system which we show naturally derives the properties of DSQs as described in the previous section. In short, the following are the main claims we argue for:

i. In all languages, a syntactic interrogative clause involves an interrogative C\textsubscript{WH}. This element enters a dependency with (one or more) *wh*-expressions in a *wh*-question.

ii. The difference between *wh*-movement and true *wh*-in-*situ* [TWhiS] languages involves the features of C\textsubscript{WH}: in a *wh*-movement language, C\textsubscript{WH} has features that require an overtly filled specifier whereas in a TWhiS language, the dependency between C\textsubscript{WH} and a *wh*-expression may be satisfied without overt movement, either by covert movement (however that is understood) or via a binding dependency.

iii. DSQs are (syntactically) declarative TPs or CPs which lack C\textsubscript{WH} and in which the *wh*-XP is in focus (position); an interrogative interpretation is derived via pragmatics.

The general architecture of our proposal is given in (20). In all languages, true syntactic questions involve a dependency (to be specified below) between C\textsubscript{WH} and a *wh*-XP. No such dependency exists in DSQs. Since there is no (movement) dependency in DSQs, the fact that DSQs are
possible in islands is expected. Furthermore, since DSQs are formally declaratives they cannot occur in contexts which syntactically select for an interrogative (i.e., an attempted DSQ such as
He asked me your boy's name is what is excluded in exactly the same way the declarative He asked me your boy’s name is Leo is).

(20) a. Syntactic interrogative

\[
\begin{align*}
\text{CP}_\text{WH} & \quad \text{TP} \\
\text{C}_\text{WH} & \quad \text{wh-XP}
\end{align*}
\]

b. DSQ

\[
\begin{align*}
(C) \quad \text{pragmatics: interrogative} \\
\text{TP} & \quad [\text{declar}] \\
[\text{wh-XP}]_{\text{FOC}}
\end{align*}
\]

The approach in (20) has some immediate consequences for the syntax of \textit{wh}-phrases. Since there is no morphological difference between moved and unmoved \textit{wh}-expressions, the null hypothesis is that \textit{wh}-phrases in regular syntactic questions and DSQs are the same elements, that is, they are equipped with the same lexical features. We are not aware of any language, for example, in which moved and unmoved (DSQ) \textit{wh}-expressions (which correspond to interrogatives) are morphologically distinct. On an account that would assign one class a feature that the other class lacked, the lack of an overt signal of this difference would be a strange coincidence. Since there is no \textit{wh}-operator in the CP in (20b), \textit{wh}-XPs must be syntactically independent of a C head; in other words, they cannot involve a feature which needs to be licensed by an interrogative C head or a feature that triggers movement of a \textit{wh}-XP. We therefore propose that \textit{wh}-phrases involve an \textit{interpretable, valued wh-feature: iQ: wh}.\footnote{We do, however, posit a featural difference between the interrogative and indefinite uses of \textit{wh}-expressions in languages like German: interrogative \textit{wh}-expressions involve an \textit{interpretable wh-feature}, whereas the indefinite ones involve an \textit{uninterpretable wh-feature} (see also Pesetsky and Torrego 2007).} This has the advantage that no other Q feature must be in the structure to license a \textit{wh}-XP, and DSQs are thus in principle possible, even in formally declarative contexts. Furthermore, being equipped with an interpretable interrogative feature, \textit{wh}-XPs have the power to contribute an interrogative interpretation by themselves. This is what we suggest is the case in DSQs. Once set in an appropriate pragmatic and semantic context, \textit{wh}-XPs can trigger an interrogative interpretation and yield a DSQ.

Regarding syntactic dependencies and feature licensing, we follow the view that the interpretability of features is independent of the notion of valuation (both interpretable and uninterpretable features can come as valued [i/uF: val] or unvalued [i/uF: __]) and that licensing is established under Agree which is valuation driven (see Pesetsky and Torrego 2007, Bošković 2009). The definition of Agree we adopt is the Reverse Agree definition given in (21) from Wurmbrand (2014).

(21) A feature F: __ on \(\alpha\) is valued by a feature F: val on \(\beta\), iff

\[
\begin{align*}
i. & \quad \beta \text{ c-commands } \alpha \text{ AND} \\
ii. & \quad \alpha \text{ is accessible to } \beta. \ [\text{accessible}: \text{not spelled-out}] \\
iii. & \quad \alpha \text{ does not value } \{\text{a feature of } \beta\}/\{\text{a feature F of } \beta\}.
\end{align*}
\]

Let us see how this system derives syntactic interrogatives. As shown in (22), interrogative C is first merged with an unvalued, but crucially \textit{interpretable} Q-feature. Given Agree as in (21), this feature needs to be valued by a c-commanding, valued Q-element, such as a \textit{wh}-phrase. This re-
requirement thus triggers movement of the *wh*-XP, which we assume is inherently valued as a *wh*-phrase (*iQ*: *wh*), to Spec,CP, a position from which it may value the unvalued *iQ*:__ of C. The result of this derivation (see (22a)) is an interrogative clause, that is, a CP typed as *iQ*: *wh*. Such a clause may subsequently be merged with a verb that selects an interrogative, such as wonder or ask, as in (22b). Merging an interrogative selecting verb such as wonder with a CP lacking *iQ*: *wh* is filtered out by however selection is implemented (see, among others, Adger 2003, Wurmbrand 2014 for feature based proposals). It is important to note that C\textsubscript{WH}, and thus the CP it heads, is unvalued until such time as the *wh*-XP moves to Spec,CP and values the *iQ*:__ of C\textsubscript{WH}. This property forces *wh*-movement to be overt in embedded interrogatives—prior to movement of the *wh*-XP, the CP is *iQ*:__ and thus not selectable by an interrogative-embedding predicate.\textsuperscript{9}

\begin{align*}
\text{(22) a. Matrix interrogative} & \quad \text{b. Embedded interrogative} \\
\text{[iQ: wh]} & \quad \text{[iQ: wh]} \\
\text{XP} & \quad \text{XP} \\
\text{C} & \quad \text{C} \\
\text{C'} & \quad \text{TP} \\
\text{TP} & \quad \text{TP}
\end{align*}

As noted above, we assume that TWhiS languages also involve a C–XP-*wh* dependency. Without delving into the many intricacies of this topic, it seems there are at least two broad families of analysis of TWhiS, plausibly reflecting two classes of phenomena or different language types (see, e.g., Cheng and Rooryck 2000, Bayer 1996, Cole and Hermon 1998). One type involves covert movement—a derivation essentially like (22), with the only difference that the *wh*-elements move covertly. There are a variety of ways of implementing this, including covert phrasal movement (Huang 1982, Pesetsky 2000), movement of an invisible feature or other sub-word element (Watanabe 1992, Pesetsky 2000, Cable 2007), or overt phrasal movement followed by pronunciation of the lower copy. For our purposes, all of these may be lumped together under the rubric ‘covert’ movement. In such a language, TWhiS should be subject to the same locality conditions as overt movement. A second strategy is (un)selective binding (see, e.g., Pesetsky 1987). In such languages/constructions, we propose that C can be inserted with a valued Q feature. If that feature binds another Q feature, the structure is interpreted as a *wh*-question. Like variable binding, unselective binding requires c-command, but is otherwise not sensitive to islands, which correctly characterizes certain TWhiS languages. Thus, while island-sensitivity is}

\textsuperscript{9} We leave open the possibility that one species of apparent *wh*-in-*situ* is derived by overt movement of the *wh*-phrase, but with the lower copy of that phrase pronounced (Bobaljik 1995). This cannot, on our view, be the derivation of DSQs, though, contra, e.g., Reglero (2007). Such a system does not discriminate between matrix and embedded interrogatives, and would incorrectly allow indirect questions without movement.
indicative of TWhiS, lack of island-sensitivity does not entail that the structure involves a DSQ—both DSQs and TWhiS established via unselective binding are possible in islands.10

Crucially however under our proposal, a DSQ does not involve a C–XP-wh dependency. For this reason—and this is the main difference between TWhiS and DSQs—DSQs may be interpreted as interrogatives, but lacking a CP marked as iQ: wh cannot be embedded by a predicate selecting a syntactically interrogative complement. Embedability is thus indicative of TWhiS. Both covert movement and unselective binding produce CPs that are embeddable, thus we need not enter into debates about the merits of individual approaches to wh-in-situ; the crucial point for us is that we have provided a formal characterization of the central difference between DSQs and TWhiS.

4.2 DSQs with wh-in focus

The languages considered in section 3 share the property that DSQs resemble simple declarative clauses in their gross word order. These are all languages in which elements in focus may remain in situ. Since we propose that DSQs are focus constructions (cf. (20b)), in languages where focus constituents occupy dedicated positions, DSQs will be characterized by a wh-expression in focus, rather than necessarily in situ. Exactly such distribution is found in Spanish, among others.

DSQs in Spanish are similar to Brazilian Portuguese, but with the additional requirement, discussed by Uribe-Etxebarria (2002), and Reglero (2007) that the non-fronted wh-expression must be final in its intonational phrase. This order appears to be distinct from wh-in-situ, in that it does not always coincide with the unmarked order of post-verbal constituents, but as Reglero discusses in detail, this requirement holds independently of post-verbal focus constituents in Spanish. For example, while the unmarked order in Spanish is DO»IO, when the DO is an unmoved wh-expression, it follows the IO, reflecting the broader generalization that focused constituents come last in their intonational phrase:

(23) [y] tú le diste a María (el) qué?
    and you CL gave to Maria the what
    ‘And you gave Maria what?’

The fact that Spanish in situ wh-expressions obey a word order generalization not evident in wh-movement languages is thus, as Reglero originally argued, not indicative of any syntactic difference about wh-in-situ among these languages, but is rather a consequence of the facts that DSQs involve wh-in-focus, and that the syntax (or perhaps prosody) of focus differs between Spanish and the languages described in section 3.

Cross-linguistic variation in the syntax of focus thus conspires to sometimes obscure the fundamental cross-linguistic similarity in the syntax of DSQs. This phenomenon is particularly important when considering DSQs in Slavic. It is widely recognized in the literature on multiple wh-fronting that not all fronted wh-expressions in Slavic undergo wh-movement—some occupy a lower, focus position (Bošković 1998). Focus movement (i.e., to a non-initial position) is possible for question formation even with a single wh-word. If these are instances of DSQs (and not binding by an interrogative C), then we expect that the non-initial fronting will, like other DSQs

---

10 There is much more to be said here. As regards putative island-insensitivity in wh-in-situ languages, there is a substantial literature debating this point (see Cheng 2003, Bayer 2006 for overviews).
conform to (1). This is correct for Polish (Lasnik and Saito 1984) and Russian (Dyakonova 2009), at least: as shown in (24) for Russian, wh-expressions in matrix questions may either undergo wh-movement or focus movement, but in an indirect question, a wh-expression must front to the initial position.11

(24) a. \{kogda\} Ivan \{kogda\} budet \{kogda\} vystupat’?
   \{when\} Ivan \{when\} will \{when\} present
   ‘When is Ivan going to present?’ [N. Radkevich, p.c.]

   b. Boris sprašivat' kogda Ivan \{*kogda\} budet \{*kogda\} vystupat’.
      Boris ask.PRES.3SG when Ivan \{*when\} will \{*when\} present
      ‘Boris is asking when Ivan is going to present.’ [Dyakonova 2009: 197]

In other words, exactly as in all the other languages examined above, it is possible to ask a question with the syntax of a declarative clause, where a wh-expression occurs in a focus position, but a clause with this syntax cannot be syntactically selected by a predicate that selects an interrogative complement. Such a predicate requires a complement that is not only interrogative in force, but is syntactically marked as such, for example, by (true) wh-movement.

Similarly for Polish, Lasnik and Saito (1984) contend that wh-expressions must occupy an A’-position at S-structure, but need not front overtly to Spec,CP (on their account, they may do so at LF). Thus, in (25a,b) the wh-word is in non-initial position in the embedded clause, but still yields a matrix question interpretation. Yet even in Polish, if an interrogative complement is selected (as by wonder), then a wh-word must be fronted to clause initial position (embedded Spec,CP), as in (25c).

(25) a. Maria powiedziała, że co Piotr kupił?
      Maria said that what Piotr bought
      ‘What did Maria say that Peter bought?’

   b. Spotkałeś mężczyznę, który jak rozwiązał zadanie?
      you.met man who how solved problem
      ‘How did you meet the man who solved the problem?’

   c. Zastanawiam się \{co\} Maria przyniesie \{co\}.
      I.wonder REFL \{what\} Maria will.bring \{what\}
      ‘I wonder what Maria will bring.’ [Lasnik and Saito 1984]

As our system now stands, we must reject the characterization of languages like Polish, English or French, as having optional wh-in-situ, or optional fronting to non-initial A’-positions. The DSQs are crucially not simply limited instances of the wh-in-situ derivation that languages like Chinese use freely. Note again that TWhiS languages such as Mandarin happily allow cases exactly like (25c) (cf. (10)), thus there is no general ban on embedded TWhiS.

Instead we assume that Polish and Russian, like English, are wh-movement languages and that cases of unmoved wh-phrases (or rather cases where the wh-elements are moved to a focus position) are instances of DSQs and not TWhiS. This approach derives the distribution in (25). That these are DSQs is further suggested by the fact that (25a,b) are environments out of which overt wh-movement would be prohibited (Polish disallows wh-extraction out of finite comple-

11 It appears there is some variation in Russian as to whether the wh-expression (and focused expressions generally) may occur in post-verbal position in (24a) (see, e.g., Dyakonova 2009).
ments), and provides a principled account for why apparently optional wh-movement languages like Polish always lose the option of wh-in-situ in embedded contexts: in syntactic questions, movement is always obligatory; the apparent optionality arises as a result of the option of DSQs. As we have surveyed briefly in section 3, the option of DSQs to form matrix interrogatives shows varying degrees of pragmatic freedom cross-linguistically, but is syntactically always restricted to matrix interrogative interpretations.

4.3 Alternatives?

In this section, we compare our feature valuation system with two other important proposals. Our assumption that wh-XPs are inserted with a valued Q-feature and C is inserted with an unvalued Q feature is identical to the feature assignment proposed in Pesetsky and Torrego [Pe&To] (2007). However, in contrast to our system, Pe&To follow an upward valuation approach to Agree. In derivations like (26), C probes downward until it finds a valued Q-feature, the uQ: wh of the wh-XP, and is then valued (upwards) by that XP. To derive movement of the wh-XP, C is equipped with an EPP feature in movement languages. The difference between movement and TWhiS languages could thus be encoded as the presence vs. absence of an EPP feature. To derive the properties of languages like Polish or BP, with (descriptively) optional wh-movement in matrix interrogatives, it would be necessary to have an optional EPP feature on matrix C, but an obligatory EPP feature on embedded Cs. While this technically yields the correct results, it raises the question of why embedded C must always involve an EPP feature in languages with an optional EPP feature on matrix C.

Furthermore, this account does not explain why of the four possible EPP combinations only three seem to be attested. If both matrix and embedded Cs lack EPP features, a TWhiS languages is derived (e.g., Mandarin); if both matrix and embedded Cs involve EPP features, an obligatory movement language is derived (e.g., English, though DSQs are then not accounted for; see also below); if matrix C involves no or an optional EPP feature, but embedded C involves an obligatory EPP feature, a language like Polish is derived. What is not attested is a language with an obligatory EPP feature on matrix C, but no or an optional EPP feature on embedded C. This would yield a language which involves obligatory movement in matrix clauses but no or optional movement in embedded clauses. Under a Pe&To feature system, there does not seem to be a
principled reason for why such languages are missing. In our account, on the other hand, this is predicted: A language is either a TWhiS language or an obligatory movement language (matrix and embedded clauses always behave alike); all languages allow DSQs, however, DSQs are only possible in matrix questions. Hence, the only ‘mismatch’ between matrix and embedded clauses that is possible is cases in which a matrix clause involves no movement (due to DSQ), but the embedded clause requires movement (due to the unavailability of embedded DSQs).

Lastly, a Pe&To feature system does not readily extend to DSQs in a language like English. Since wh-XPs are specified as uQ: wh, they do not carry interrogative force and thus may not be able to license an interrogative interpretation by themselves. A possible solution may be to assume a special DSQ C which licenses the uninterpretable Q-feature of wh-XPs in DSQ contexts but by a means different from Agree. In this regard, Pe&To’s approach would become very similar to the proposal of Pires and Taylor (2007), where it is explicitly argued that DSQs must be licensed by a special Q operator. The following discussion thus applies to both proposals and similar ones involving a syntactic licensing requirement of wh-XPs in DSQ contexts (see for instance Cheng and Rooryck 2000, Zocca DeRoma 2011, among others). Pires and Taylor (2007) suggest that DSQs in English are cases of wh-in-situ, specifically, wh-in-situ involving the strategy of unselective binding. As mentioned above, the main hurdle for such proposals is to block DSQs in embedded questions. Since TWhiS languages readily allow indirect questions with wh-in-situ, it is hard to see how the special DSQ C could be blocked from being selected, in a principled way. Why, for instance, can’t this special C be selected (it can in TWhiS languages)? Why is it that all languages, even ones that most persistently have been described as obligatory wh-movement languages (such as English) allow DSQs? If the special DSQ C is a lexical item, we would expect it to be absent in some languages. However, once we control for pragmatic factors, DSQs seem to be available cross-linguistically, exactly as expected under our approach where wh-XPs involve an interpretable Q-feature which licenses an interrogative interpretation even in the absence of an interrogative syntax. In sum, in approaches involving (or requiring) licensing of wh-XPs, the unembedability of DSQs becomes an accidental, idiosyncratic restriction arising only in (obligatory and optional) wh-movement languages, whereas it follows on principled grounds from our approach.

4.4 A final note—French

Up to this point, we have put aside discussion of French, although it is one of the most widely discussed cases of ‘optional’ wh-in-situ in a wh-movement language. French is somewhat of a two-edged sword, in part as there are competing descriptions of the facts, and evidently variation among speakers. All analysts agree that French respects the basic contrast, consistent with the generalization in (1). Wh-in-situ is freely available to form non-echo questions in simple clauses, as in (27a), but such questions are strongly ungrammatical as an indirect question; wh-movement in the embedded clause is obligatory as in (27b,c).

---

12 Compare verb movement (to C0), which across (most of) Germanic has exactly the property of being obligatory in matrix questions but impossible in indirect questions.

13 Pires & Taylor also propose that the pragmatic restrictions on DSQs are attributable to (in some way or another) the presence of this special complementizer. But see Tieu (2011) for critical remarks.
Questions with Declarative Syntax

(27) a. Il parle de quoi?
   He talks of what
   ‘What is he talking about?’
b. *Je me demande tu parles de quoi.
   I me ask you talk of what
   ‘I wonder what you are talking about.’
c. Je me demande de quoi tu parles.
   I me ask of what you talk
   ‘I wonder what you are talking about.’ [Obenauer 1994: 291-2; P. Schlenker, p.c.]

Beyond these poles, the French are disagreeable. Broadly speaking, there are at least three different characterizations of the data. An important starting point is Obenauer (1994) who offers an extended comparison of French and English. Under Obenauer’s characterization (see also Pollock 1998, Starke 2001), French is like English in allowing wh-in-situ in embedded (non-interrogative) clauses as in (28), and even in (weak) islands as in (29) and (30) (see also Starke 2001, Shlonsky 2012).

(28) a. Tu veux que je le fasse quand?
   you want that I it do when
   ‘You want me to do it when?’ [Obenauer 1994: 319]
b. Tu crois que Jean a acheté quel livre?
   you believe that Jean has bought which book
   ‘You believe that Jean bought which book?’ [Pollock 1998: 189]

(29) a. Il s’est defendu [en accusant qui]?
   he refl-is defended [by accusing who]
   ‘He defended himself by accusing who?’
b. *Qui s’est-il defendu [en accusant t]?
   who refl-is-he defended [by accusing t] [Obenauer 1994: 296]

(30) a. Vous connaissez des gens qui pourraient héberger combien de personnes?
   you know of people who could host how many people
   ‘You know people who could host how many people?’ [Obenauer 1994: 297]
b. *Combien de personnes connaissez-vous des gens qui pourraient héberger?
   how many people know-you of people who could host

However, unlike English, French on this characterization does obey various types of apparently syntactic locality conditions. On Obenauer’s characterization, embedding the islands in (29)-(30) still further yields unacceptability; later authors (Starke 2001, Shlonsky 2012) characterize this as a weak-strong island asymmetry. Many authors also contend that various quantificational elements, including negation, occurring above the in-situ wh-expression lead to degradation.

Chang (1997) (and following her Bošković 2000 and others) presents a much more restrictive variety than Obenauer (1994) and Starke (2001). In this variety, while the contrast in (27) obtains, wh-in-situ is significantly degraded in embedded clauses generally—these authors mark sentences like (28) as unacceptable. Thus while the generalization in (1) holds, it does for rather trivial reasons, as a special case of a broader generalization.
In the other direction, Starke (2001) presents a more permissive, colloquial variety of French than Obenauer. Starke’s characterization is essentially similar to Obenauer’s, noting the strong vs. weak island asymmetry, and also the intervention effect of elements such as negation, but Starke contends that these do not yield unacceptability per se. Rather, these sentences are limited in their contexts and intonation patterns; but when these variables are controlled for, wh-in-situ is seen to be acceptable even in strong islands and under negation and other putative interveners (Starke notes carefully that the relevant contexts are nevertheless distinct from echo questions).

We have neither the space nor the resources to sort out the French situation in the remaining pages of this short paper. We note on the one hand that all varieties of French thus far reported are consistent with the generalization in (1)—French thus supports our main contention that the lack of movement and unembeddability are intimately entwined. On the other hand, we note that most authors, cutting across the three varieties reported, argue for LF-movement of wh-in-situ and thus a covert version of the dependency with C that we reject on the basis of languages like English.

Of the competing accounts, Bošković (2000) analysis stands out in offering a proposal that restricts the construction to matrix questions. In brief, under Bošković’s account (like those of its competitors), the French interrogative complementizer C_{WH} always triggers movement, but C_{WH} may be inserted into the derivation early, in which case movement is overt, or late (at LF), in which case the movement is covert (yielding apparent wh-in-situ). Given other assumptions of the framework in which Bošković (2000) is couched, the covert movement option is restricted to matrix clauses: because of the cyclicity of derivations, in order for C_{WH} to be inserted in an embedded clause it would need to be inserted overtly, prior to the merge of the embedded clause as complement to the matrix predicate. In this way, it is only in matrix questions that insertion of C_{WH}, and thus movement of the wh-expression, may be delayed until LF.

While Bošković’s account thus meets the general desideratum we have set forth, relating the in-situ property to the inability to be selected, Bošković treats the numerous restrictions of the narrow variety of French as key support for his proposal: the inability of the wh-in-situ to be beneath negation or quantifiers, or to occur in embedded clauses at all, are taken as key evidence for the specific account he sets forth. Thus, as other authors have noted (Reglero 2007, Wood 2009, Zocca DeRoma 2011), the properties that support Bošković’s account of the variety he describes keep that account from generalizing to languages that fail to show a general ban on embedding, that allow wh-in-situ under ‘interveners’ such as negation, and which allow wh-in-situ in islands. Thus, while Bošković’s proposal provides an account of the language particular properties of one variety of French, it does not generalize. Conversely, an account such as ours, which focuses on the general property of DSQs, does so at the expense of the language-particular properties of various languages.\footnote{Dyakonova (2009) claims that Russian patterns with the restrictive variety of French, but no other language we have investigated does. There are additional differences among individual languages which are not captured by our account. For example, Bayer (2006) notes that in German warum ‘why’ resists forming DSQ contexts, contrasting with the roughly synonymous aus welchem Grund ‘for what reason’.

8 Conclusion

Although we have left several loose ends, the landscape of DSQs can be characterized as follows. In all languages that we have investigated with classical wh-movement (i.e. to the leftmost periphery of the clause), an ‘optional’ interrogative strategy exists—even in non-echo contexts—
in which the clause has declarative syntax, with a \textit{wh}-expression in focus position (i.e., \textit{in-situ} where there is no special focus syntax). There is variation (both cross-linguistic and intra-speaker) in the pragmatic contexts in which this strategy is available, and, most strikingly in French, also in the syntactic contexts that tolerate DSQs. Despite this variation, there is one point of absolute stability: the \textit{in-situ} construction is blocked in selected questions. The very systematicity of the last point constitutes an argument that ‘optional’ \textit{wh-in-situ} in \textit{wh}-movement languages should not be assimilated to the \textit{wh-in-situ} strategy of TWhiS languages, which lack such a restriction. This property is theoretically important in that it speaks to the nature of what is selected; but it is also of interest in that it constitutes the one island of stability in a sea of other cross-linguistic variation in both the clause-internal syntax and overall distribution of DSQs. We have sketched a feature-based account that weds the unmoved nature of the \textit{wh}-expression to its unembeddability, relating this centrally to the notion of syntactic selection set out in \textit{Aspects}.

Finally, our approach leads us to expect that DSQs should be possible in principle in all languages with \textit{wh}-expressions. In a TWhiS language, these will be very hard to distinguish from \textit{wh-in-situ}. But not impossible. Under our approach, DSQs and TWhiS should both be possible as matrix questions, but DSQs should be impossible as indirect questions. To the extent there are TWhiS languages that show island effects, we thus predict a class of languages in which, as a matter of observation, island effects with \textit{in-situ} \textit{wh}-expressions are avoided in matrix question interpretations (which may be DSQs) but create violations when the intended interpretation is an indirect question (where DSQs are impossible). We do not know at this time whether such a class of languages exist, but leave this as an open conjecture.

References

Bošković, Željko. 2009. Unifying First and Last Conjunct Agreement. \textit{Natural Language and Linguistic Theory} 27.3:455-496.


THE SUBTLE DEPENDENCY BETWEEN COMPETENCE AND PERFORMANCE

CRISTIANO CHESI
ANDREA MORO
Center for Neurocognition, Epistemology and theoretical Syntax (NEtS), IUSS Pavia

1 Introduction to the classic competence vs. performance divide

Since Chomsky’s (1957) seminal work, we think of a grammar as a formal device (G) providing an explicit description of all productive constraints (e.g. rewriting rules) that restrict the possible geometry of all the significant relations among words (or morphemes) and other (intended) syntagmatic categories. A constituency-based graph (i.e. a syntactic tree) represents the result of the application of the relevant rules and principles and indicates, unambiguously, both first order relations, like precedence and dominance among (certain) nodes, as well as second order relations like (anti-)symmetric C-command. The primary goal of G, for a given language L, is to distinguish “grammatical” sequences of words from “ungrammatical” ones, where L is meant to be an infinite set of word sequences each associated with (at least) one relevant structural description. In mentalistic terms, G expresses the linguistic competence, that is “speaker/hearer’s knowledge of his/her language” (Chomsky 1965:4).

G is descriptively adequate with respect to L, if and only if it succeeds in generating nothing but the infinite set of sentences that are accepted by the native speakers of L and if, in addition to the exact definition of the L boundaries, it also suggests how the accepted sentences are understood by an “ideal speaker-hearer” (Chomsky 1965:5). In this sense, G is a theory of L (Chomsky 1965:24) since it describes all possible sentences in a given corpus but also any new possible sentence that is not (yet) in any given finite collection of well-formed expressions.

Notice that any grammatical formalism G must be used “mechanically”, that is, given an adequate algorithm (potentially included in G itself), we should always be able to decide if a sentence is well-formed or not without any ambiguity and without any other external knowledge. Moreover, in order to be computable, the algorithm should attain to this goal in a finite amount of time, that is, the computation should terminate in a finite number of steps.

This clear picture might get blurred by the fact that, given a phrase structure grammar G, many different algorithms can perform the generation (and the recognition) task by making use
of the constraints expressed by $G$: rewriting rules of the context-free kind\(^1\), for instance, are
descriptive constraints that can be used by different parsing algorithms in order to recognize\(^2\) the
structure of a well-formed sentence in a finite amount of steps (see Earley or CKY algorithms,
Jurafsky and Martin 2009). The compatibility of $G$ with many algorithms might be problematic
since it might well be the case that for a given sentence $s$, different algorithms on $G$ could take
more or less time, or, in certain cases, not even terminate (§2). In more precise terms, different
algorithms for $G$ can have different complexity, where complexity can be precisely expressed in
terms of time (number of steps) and space (memory needed to store working information) as
discussed by Barton, Berwick and Ristad (1987).

A problem arises if we assume that the human linguistic performance, that is the actual use of
the linguistic competence (Chomsky 1965:10) taking into consideration time and space limits, is
an approximation of the behavior of a specific algorithm $A$ applied to $G$. On the one hand,
assuming some restriction on $A$ is the only way to justify a divergence from the intuition of the
real native speaker ($RS$) and the ideal speaker ($IS$); on the other, since many algorithms $A$ can be
applied to $G$ and many grammars $Gs$ can circumscribe similar set of expressions, both including
or excluding certain critical sentences for which the actual and the ideal native speaker intuition
diverges, we are left with the ambiguity of deciding whether $A$, $G$ or both are inadequate.

In fact, we can find examples of two different kinds of divergence between $RS$ and $IS$: on one
side, there are sentences that should be well-formed, according to $G$, but that are considered ill-
formed according to $RS$ (e.g. self-embedding, (1), Chomsky 1965:11, Miller & Chomsky 1963),
on the other, there are expressions that should be ill-formed but that are processed consistently
by real speakers, most of the time being accepted, like (2) (i.e. subject islands violation with
pied-piping, Bianchi & Chesi 2014, in press):

(1) [The man [who the boy [who the students recognized] pointed out] is a friend of mine]

(2) [Of which masterpiece], is [one reproduction _i_] already available?

Since, for the great majority of cases, $IS$ and $RS$ intuitions seem to match, the situation is less
disturbing than expected. There are however extreme cases, as usually in science, as (1) and (2),
that are richly informative about the competence vs. performance divide and that ask for an
explanation. In this paper, we suggest that the tension related to this divide did not disappear
within the last 50 years (indeed, it has been increased by the advent of the Minimalist Program),
then we want to dig into some crucial evidence, following Chomsky’s original intuition, namely
that there is no a serious performance model that is not “competence-based” (Chomsky 1965:10).
In the end we will conclude that a linguistic theory that subsumes the real speaker performance
in contexts like (1) and (2) is explanatory more adequate than a theory that relegates (1) to
performance facts and (2) to exceptional cases.

---

\(^1\) Context-Free Grammars (CFGs) are Type 2 phrase structure grammars in the Chomsky hierarchy. CFGs use
rewriting rules of the form “$A \rightarrow \gamma$” where “$A$” is a non-terminal symbol and “$\gamma$” represents any sequence of
terminal or non-terminal symbols (included the null one) and “$\rightarrow$” is the rewriting symbol expressing immediate
dominance: only one non-terminal symbol is allowed to the left side of the rewriting symbols in CFGs and anything
is possible to the right side.

\(^2\) In psycholinguistic terms, parsing and production refer to human performance tasks; in recursion and automata
theory, the corresponding tasks are usually called recognition and generation respectively. Here we will use parsing
and recognition, as well as production and generation as synonyms, both indicating mechanical procedures.
In order to discuss these points, the paper is organized as follows: first, we will state in a formally clear way the “competence paradox”, that the complex relation between algorithmic and constraining part of RS and IS (§2); then we will present some relevant data on self-embedding and extraction from subject islands, pointing out to new empirical evidence not yet available in the ’60s (§3). It will be clear in the end (§4) that Minimalism (Chomsky 1995) exacerbates the competence paradox, since constraints on the possible algorithmic procedures are provided that are not easily compatible with any performance task. In the end, we will sketch few alternatives that attempt at bridging the gap between RS and IS, suggesting configurational triggers for non-local relations (on the line of Moro 2000) and a different orientation of the structure building operations (Chesi 2014).

2 The competence paradox

As long as the infinite set of well-formed sentences is independently and precisely defined (that is, there is at least an ideal speakers that consistently generates and recognizes all well-formed sentences on the one hand and rejects ill-formed ones on the other for a given language) the “competence problem” is well stated. Unfortunately, there is no ideal speaker’s competence of L that is distinguishable by A(G), where A is an algorithmic procedure taking as argument the grammar G, and both A and G can be independently tuned in order to generate or recognize L³.

Let us exemplify this with a simple, non-trivial (namely recursive) CFG that generates any possible sequence of (non-null) a s and b s in this order (S, as usual, is the starting symbol; rules are numbered for convenience):

(1) \[ G = \]

1. \( S \rightarrow aSX \)
2. \( S \rightarrow aX \)
3. \( X \rightarrow Xb \)
4. \( X \rightarrow b \)

Rewriting rules “\( \rightarrow \)” can be read bottom-up (from the right side of the rewriting symbol to its left side) or top-down (in the opposite way, that is the standard sense of the rewriting symbol). This already permits two different kinds of generation algorithms based on distinct problem spaces⁴ that are the result of different order of application of the possible rules; below an example of the problem space generated by a top-down, left-right expansion of non-terminals (left-most symbol is expanded first):

(2) Partial top-down, left-right problem space expansion (the states of the problem space are labeled with the result of the substitution; the labels on the arrows leading to a certain state indicate the rule applied; black boxes are the generated terminal strings that end the derivation)

---

³ One might object that \( A \) can be included in \( G \). This possibility (exploited in Minimalism) is not logically distinguishable from the one we discuss here, since if there is an algorithmic part in \( G \), independent of the constraints/principles/rules expressed by \( G \), we can always consider this as \( A \), hence unbundle the algorithm from \( G \) (e.g. separating structure building operations, \( A \), from lexical features, \( G \), just to use familiar minimalist terms).

⁴ The problem space represents all possible reachable states that can be obtained by applying legal rules.
Assume that the algorithms we want to test only deal with serial expansions\(^5\) and that we want to compare how “complex” is generating the string “ab” with these different algorithms, all using the grammar in (1). Below the result of the relevant exploration of the problem spaces generated using 5 different algorithms:

\[\text{(3) A1. top-down, left-right, breath-first (short-rule first)}^6:\]

\[\text{goal: } ab; \text{ computation: } S, aX, aSX, ab \quad \text{(it terminates in 3 steps)}\]

\[\text{A2. top-down, left-right, breath-first (long-rule first):}\]

\[\text{goal: } ab; \text{ computation: } S, aSX, aX, aaSX, aaXX, ab \quad \text{(it terminates in 6 steps)}\]

\[\text{A3. top-down, left-right, depth-first (long-rule first):}\]

\[\text{goal: } ab; \text{ computation: } aSX, aaSX, aaSXX... \quad \text{(it never terminates!)}\]

\[\text{A4. Bottom-up, left-right, depth-first (greedy, short-rule first)}^7:\]

\[\text{goal: } S; \text{ computation: } ab, Xb, SXb, X \quad \text{(solution not found!)}\]

\[\text{A5. Bottom-up, right-left, greedy (short-first):}\]

\[\text{goal: } S; \text{ computation: } ab, aX, aXb, S \quad \text{(solution found in 4 steps)}\]

As it is clear, considering the (simpler) generation problem, there are many non-equivalent algorithms all using \(G\). Their non-equivalence is demonstrated by the fact that some algorithm does not even terminate (\(A3\)) while others might conclude the computation without finding a solution (\(A4\)); their different complexity, on the other hand, is suggested by the fact that \(A1\) finds the relevant string in 3 steps, \(A2\) in 6 steps and \(A5\) in 4 steps (as we said, time complexity is

\(^5\) In case of multiple possible application of different rules the algorithm expands one rule per time. This is the case of rule 1. and 2., for instance: reading these rules top-down, they both can be applied whenever \(S\) is found. This produces branching as indicated in (2) (first bifurcation after the root state, marked as “\(S\”)).

\(^6\) Breath-first is opposed to depth-first: in the breath-first case, all alternative expansions are evaluated before going deeper; in the depth-first case, one expansion is pursued up to a terminal string is rewritten.

\(^7\) “Greedy” means exploring first the rules with the maximum number of matches: if “ab” is the sequence to match, and there are two rules compatible with the left-prefix, “\(X \rightarrow ab\)” and “\(Y \rightarrow a\)”, \(X\) will take over.
measurable also in terms of number of steps). This means that if we want to simulate a performance algorithm, we have many different alternatives that produce different (sometimes conflicting) outcomes.

From a generative point of view, any algorithm that explores the problem space completely would simulate the IS competence (e.g. A1 or A2), but if we want to explain the possible mismatch between RS and IS, there are just few possible solutions that we could consider:

(4) a. Algorithmic Independence
IS and RS use different algorithms to explore the problem space generated from G;
b. Algorithmic Equivalence
IS and RS use the same algorithm A, but RS is affected by complexity (e.g. longer derivations are harder to be computed);
c. Algorithmic Equivalence with independent constraints sensitivity
IS and RS use the same algorithm, but RS is sensitive to some property of the constraints applied and IS to others (or none);

In the first case (Algorithm Independence) $A_i$ and $A_r$ (i.e. the algorithms used by IS and RS respectively) are different and it is possible to expect that for some $s \in L$, $s$ is generable by $A_i$ and not by $A_r$ (or the way around); though logically possible, this solution is highly unexplanatory and requires the formulation of two unrelated procedures; moreover it can not guarantee whether $A_i$ or $G$ are inadequate: in order to account for (2), we should first guarantee that (2) is generable given an appropriate $G$, then that $A_i$ prevents (2) from being generated while $A_r$ generates it, in the end, giving the suspicion that $A_i$ is simply inadequate. The hypothesis (4).b (Algorithmic Equivalence) is somehow more explanatory since just one algorithm needs to be stated ($A_i = A_r$), but it is a matter of fact that its complexity is the only restriction affecting $A_r$; in fact, it is impossible to explain cases in which a sentence $s$, such that $s \not\in L$, is generable (and recognizable) by $A_r$ and not by $A_i$. Under this circumstance, the language $L_r$ generable by $A_r$ should be fully included in $L$ (namely by the language generable by $A_i$). Last hypothesis, (4).c (Algorithmic Equivalence with constraints sensitivity), could predict that certain rules chain, like expansions of the same recursive rule in sequence, is penalized by $A_r$ that simply confounds similar rules in self-embedding cases (much on the line of Miller & Chomsky 1963), but how to explain (2)? In the end, all these options could predict a divergence between IS and RS. Notice however that in all cases, we are left in a fideistic condition in which the adequacy of $G$ (as stated in §1) is guaranteed only by the grammar itself. This leads to the competence “paradox”:

(5) The competence “paradox”
i. Given $A_i$ and $A_r$, two distinct algorithms (inspired by the ideal and the real native speakers of $L$, respectively), a grammar $G$ is adequate with respect to $L$, if and only if $A_i(G)$ generates nothing but any possible sentence $s \in L$ where the relevant set $L$ is independently generated by $RS$ (the real native speaker of L).
ii. If RS and $A_i(G)$ diverge, $G$ is adequate if only if there is $A_r$ such that $A_r(G)$ generates nothing but any possible sentence generable by $RS$.
iii. If $A_r$ exists, there is no reason to assume $A_i$. If $A_r$ does not exist $A_i(G)$ (hence $G$) is not adequate.

---

It would be interesting to try to demonstrate whether or not an algorithm can uniformly find a solution quicker than other algorithms (see Russel & Norvig 2003:91), but this goes beyond the scope of this paper.
This is a “paradox” since we concluded that there is not theory of competence if there is no theory of performance! Indeed, we think we can make sense out of it: at the beginning of ‘900, the physicists gave birth to quantum theory on the basis of a wave theory of light built on controversial experimental observations (Mehra and Rechenberg, 1982). Taking inspiration from this, we should not relegate systematic RS data to the “performance” rubric, but integrating them, as best as possible, within our theory of $L$. This could lead to less paradoxical conclusions.

3 Rediscovering (and refining) the same old data

Since Aspects, the empirical evidence greatly increased in the last decades, both in terms of numbers and quality of sources. In this section we will briefly review the main evidence bearing on the two cases we want to discuss: self-embedding and island domains. In the first case complexity signatures (including brain imaging studies) will be used to show that non-local dependencies are sensitive to feature-matching in a non-trivial way (§3.1); in the second case, we will show that we can deep our understanding using very subtle oppositions by teasing apart variables whose correlation is only evident using robust experimental methods (§3.2).

3.1 New evidence on self-embedding and intervention

Self-embedding is a special case of a nested construction in which two constituents, A and B, with B self-embedded in A, are of the same kind and B falls totally within A (that is, A dominates B and there is some non-null element of A to both the right and to the left side of B, Chomsky 1965:12). The example in (1), repeated below for convenience, is exactly a case of self-embedding, since one relative clause (RC) is fully included in the other, while the subject and the verb of the superordinate RC respectively precedes and follows the whole inner RC:

(1) ??[The man [RC who the boy [RC who the students recognized] pointed out] is …]

Speculating on the unacceptability (which is a matter of performance, as opposed to ungrammaticality, which is a matter of competence) of this configuration, Chomsky mentions to “suggestive observations concerning limitations on performance imposed by organization of memory and bounds on memory” (Chomsky 1965:10). He also concludes that “if [the observation that self-embedding contributes even more radically than nesting to unacceptability] is correct, then we have evidence for a conclusion about organization of memory that goes beyond the triviality that it must be finite in size” (Chomsky 1965:14). This is because we should not expect an asymmetry in acceptability between standard nesting and self-embedding if memory size limits would be the only parameter to be tuned to explain RS and IS asymmetry.

As a matter of fact, the presence of similar asymmetries has been deeply studied in RCs and, for instance, the opposition between Subject (SR, (6).a) vs. Object Relative clauses (OR, (6).b) has been widely documented:

(6) a. The boy [who recognized the students] is a friend of mine
   b. The boy [who the students recognized] is a friend of mine
The contrast between SRs and ORs, with SRs easier to be generated (and understood) than ORs, is very solid and has been reported using many different approaches: King and Just (1991) started a long tradition of self-paced reading experiments and described a significant slowdown in the critical regions: the RC subject and verb, in ORs, compared to verb and object in the SR condition, revealed longer reading times. These results match with the evidence revealed using standard complexity measurements in eye-tracking experiments (e.g. first-pass and first-pass regressions within the relative clause region⁹, Traxler et al. 2002). The same contrast has been investigated also using brain activity analysis: Just et al. (1996) revealed an increase in the volume of neural tissue activation (number of voxels, i.e. volumetric pixels, produced by an fMRI imaging system¹⁰), mainly in language centers (both on Broca’s and Wernicke’s areas), when ORs are processed (e.g. “The reporter, [that the senator attacked _,] admitted the error”) with respect to SRs condition (e.g. “The reporter, [that _ attacked the senator] admitted the error”); both SR and OR show increased activation with respect to a baseline activation produced by two coordinated active sentences (e.g. “The reporter attacked the senator and admitted the error”). Their interpretation of this fact relies on a left-right processing assumption: the coordinated active sentences are uninterrupted, since the DPs feed the required subject positions in both conjoined active sentences, consistently with canonical SVO order. On the other hand, SRs are more complex than coordinated structures since the matrix sentence (in the example above “The reporter admitted the error”) is ‘interrupted’ by the RC (i.e. “the reporter, [that _ attacked the senator]”) leading to embedding, in the sense discussed above. Here the relative clause head (“(the) reporter” in the example¹¹) feeds the subject position both in the matrix clause and in the relative clause. This does not happen in ORs, where the RC has its own subject, (“the senator”, in the example) which intervenes between the RC head and the gap in the object position within the ORs (the reporter, [that the senator attacked _]).

The debate on such asymmetry source has been reactivated recently: Friedmann et al. (2009) propose that the greater processing complexity of OR with respect to SR might be due to purely grammatical facts. Their solution goes in the direction that formal constraints, like the locality principle (Rizzi 1990), which are part of our competence, if properly interpreted, can lead to an explanation of performance effects. In the specific case, when a non-local relation is established between X (the head of a non-local dependency) and Y (the tail of such dependency) and a structural intervener Z is present in between, processing difficulty is assumed to be proportional to the number and kind of relevant features shared between the moved item and the intervener. In their work, three main cases are discussed: the identity case, in which X and Z share the crucial features triggering the dependency (e.g. [+A X], [+A Z]); the inclusion case, in which Z has only a subset of the features of X needed to trigger the dependency (e.g. [+A +B X] and [+A Z]) and the disjunction case, in which X and Z are fully distinguishable and the relevant features on X triggering the dependency are not present in Z (e.g. [+A X], [+B Z]). The identity case leads to full

⁹ “First-pass” is the sum of all the fixations beginning with the reader’s first fixation of a region until the reader’s gaze leaves such region (i.e. the relative clause). “First-pass regressions” counts how many times the reader’s gaze crosses the left edge of a relevant region (e.g. the verb segment within the RC) after a first-pass fixation.

¹⁰ Functional Magnetic Resonance Imaging (fMRI) is an imaging technique that allows us to visualize neural activity in vivo by detecting associated changes in blood flow by using blood-oxygen-level-dependent (BOLD) contrast detectors. fMRI has an excellent spatial resolution and a decent temporal resolution.

¹¹ We have no space here to discuss different interpretations of the relation between the head and the gap in restrictive RC. However both raising and matching analyses require a non-local relation between the head and the gap to be established (see Bianchi 2002 for a review). See also Friederici (2012) for a comprehensive review on the neurological evidence on RCs processing.
ungrammaticality (as in the classic wh-islands effects (e.g. “\([+_\text{wh} \text{ what}]_X \text{ do you wonder } [+_\text{wh} \text{ who}]_Z \text{ read } \_X\)”), while inclusion and disjunction are generally accepted with inclusion (e.g. “\([+_\text{wh} \text{ +NP what}]_X \text{ do you wonder } [+_\text{wh} \text{ who}]_Z \text{ read } \_X\)”) harder to be processed that disjunction (e.g. “\([+_\text{wh} \text{ what}]_X \text{ do you wonder } [+_\text{NP John}]_Z \text{ read } \_X\)”). This assumption seems to be borne out since, exactly in the case of restrictive relative clause, ORs (and not SRs) create an inclusion context, because of the presence of the overt subject.

The theory is fine grained enough also to account for subtle differences in the featural make-up of the constituents involved in the dependency: studying object cleft sentences, by means of self-paced reading experiments, Warren and Gibson (2005) revealed a scale of complexity dependent on the DPs occupying the head and the subject of the cleft sentence, with DPs type varying from definite descriptions to proper names or pronouns. The rough results of their experiments assessing all possible DP-type combinations are reported below (Warren p.c.):

<table>
<thead>
<tr>
<th>Sentence</th>
<th>reading time (ms) at avoided (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. It was the banker that the lawyer avoided _ at the party</td>
<td>365 (19)</td>
</tr>
<tr>
<td>b. It was the banker that Dan avoided _ at the party</td>
<td>319 (12)</td>
</tr>
<tr>
<td>c. It was the banker that we avoided _ at the party</td>
<td>306 (14)</td>
</tr>
<tr>
<td>d. It was Patricia that the lawyer avoided _ at the party</td>
<td>348 (18)</td>
</tr>
<tr>
<td>e. It was Patricia that Dan avoided _ at the party</td>
<td>347 (21)</td>
</tr>
<tr>
<td>f. It was Patricia that we avoided _ at the party</td>
<td>291 (14)</td>
</tr>
<tr>
<td>g. It was you that the lawyer avoided _ at the party</td>
<td>348 (18)</td>
</tr>
<tr>
<td>h. It was you that Dan avoided _ at the party</td>
<td>311 (15)</td>
</tr>
<tr>
<td>i. It was you that we avoided _ at the party</td>
<td>291 (13)</td>
</tr>
</tbody>
</table>

Warren and Gibson explain the fact that the full DPs matching condition (7).a was harder than (7).b (full DP - proper noun mismatch condition) or (7).i (pro matching condition) under a memory load account, since inserting a definite DP has an integration cost that is greater than the cost of integrating a proper noun or a pronoun in the structure; this is because proper nouns and pronouns are lower on the referentiality hierarchy scale (Warren & Gibson 2005). Their account however falls short in explaining why the proper noun matching condition (7).e registers reading times on the critical verbal region that are comparable with the definite description matching condition (7).a. Due to the accessibility hierarchy scale, proper nouns should have been uniformly more accessible (hence faster to process) than definite DPs.

On the other hand, embracing a feature-based perspective, where definite descriptions constituents decorated by \([+_\text{D} \text{ N}]\) features, proper nouns by \([+_\text{D} \text{ N}_{\text{proper}}]\) features (with N and N_{\text{proper}} distinguishing the “lexical restriction”, on the line of Belletti and Rizzi 2013) and finally pro with just \([+_\text{D}]\) features (because of the missing lexical restriction), we would expect definite description and proper noun matching conditions to be harder than the pro matching condition as

---

12 Here we simplified a bit their points in order to keep the discussion focused on few relevant contrasts. Moreover, we do not have space to discuss other similarity-based accounts as well as “usage-based” models. See Belletti and Rizzi (2013) for a review.
well as other mismatch condition since in the number of feature shared is greater in (7).a and (7).e than in the other cases\textsuperscript{13}.

In conclusion, we observe that experimental evidence suggested a gradable description of the RS performance and that this might be explained in a promising way looking at formal constraints that characterize also IS competence, that is, the sensitivity to locality constraints expressed in terms of relevant features. This seems to confirm Chomsky’s intuition that “memory” (where “memory” is naively consider any kind of formal device that would make possible a non-local dependency) is sensitive to precise aspects of the phrase structure. Indeed, here we realized that “bracketing”, as in embedding, plus “labeling of bracketing” (Chomsky 1965:10) might not be enough to explain these facts; accessing a deeper feature structure, as well as taking position on the directionality of the algorithm used might be necessary.

3.2 New evidence on Subject Islands
On the other side of our grammatical boundaries, there are cases that create the converse problem of self-embeddings: while in §3.1 we dealt with sentences that should be generable by IS but that are hard (or impossible) to be processed by RS, here we will focus on sentences that can be correctly processed by RS but that are usually supposed not to be generable by IS. This is the case of subextraction from certain subject islands (example (2) in §1 repeated below):

(2) [Of which masterpiece]i is [one reproduction \_i ] already available?

Notice that the subject constituent in the preverbal position is a (simple) case of left-branching construction in the sense of Chomsky (1965:12): Chomsky notices that left-recursion reduces acceptability; even though he focused on genitive constructions like “[[[John]’s brother]’s father]’s uncle]” (1965:13), also right-recursion on the left-branch of the tree (i.e. ROs formation on preverbal subjects) creates self-embedding (this might suggest that the cases discussed in §3.1 and the one we will discuss here might not be so unrelated).

Our argument here hinges on the fact that the idealization of subject islands, considered as absolute domain from which subextraction is always banned (Huang 1982), has been seriously challenged both from a grammatical and from a psycholinguistic point of view recently. From the grammatical perspective, Chomsky himself (contradicting Barriers assumptions, Chomsky 1986:31\textsuperscript{14}) observes that certain subjects allows for subextraction (Chomsky 2008:147):

\textsuperscript{13} Belletti and Rizzi 2013 assume that the head of the cleft is marked with a +R feature that distinguishes the cleft head from the subject. This permits to consider these contrasts under the inclusion case discussed above. See Chesi and Moro (2014) and Chesi (2014) for precise complexity metrics on these data.

\textsuperscript{14} Chomsky reported an elusive contrast in subextraction with respect to preposition stranding. On extraction from adjuncts, preposition stranding seems to be favored (Adriana Belletti’s original observation):

(i) a. *he is the person to whom [they left [before speaking \_ ]]  
   b. *he is the person who [they left [before speaking to \_ ]] (before meeting \_)

While on subject subextraction the opposite seems to be true:

(ii) a. ?he is the person of whom [pictures \_ are on the table]  
   b. ?he is the person who [pictures of \_ are on the table]

Chomsky decided to “put these matters aside here, merely noting a potential problem” (Chomsky 1986:32); see Bianchi and Chesi (in press) for a proposal of explanation of the contrast in (ii).
(8) a. *[Of which car], did [the (driver, picture)] _i cause a scandal?
   b. [Of which car], was [the (driver, picture)] _i awarded a prize?

Unless we assume that between the PP and the gap there is no a real dependency, we should enforce our grammatical power in order to include selectively (4).b and keep out (4).a. Bianchi and Chesi (2014) reviewed different proposals aiming at explaining the intricacies within the exceptional cases of subject island subextraction, eventually concluding that:

i. Those are real case of “sub-extraction” and on the dependency formation constraints active in (4) must be explained (pace Longobardi 1991 and Cinque 1980); ii. The oppositions between base vs. derived subject position (Takahashi 1994) as well as internal vs. external argument (Chomsky 2008), per se, are not sufficient to explain the contrast revealed in (4); iii. The sole possible explanation is considering the interpretation of the subject (on the line of Diesing 1992), distinguishing presuppositional subjects (i.e. semantic categorial structures, in the sense of Ladusaw 1994, in which the subject is interpreted outside the predicative nucleus) from non-presuppositional ones (i.e. thetic structures, Ladusaw 1994, in which the subject must be interpreted within the predicative nucleus); from the first, extraction is banned; from the second, that are interpreted in the “reconstruction” (base-generated) site, extraction is possible.

The conclusion that semantic interpretative options must be considered during the derivation does not weaken the main assumption on competence but simply suggests that also interpretative constraints must be taken into consideration by the algorithmic part of the grammatical knowledge. In fact, assuming standard brick over brick (that is merge-based) phrase structure building we are left in an unsatisfactory condition: we should decide whether a certain subject will be interpreted in the base position or not (or, even worse, undo an already occurred movement operation forcing “reconstruction” in the base position) on the basis of part of the predicative structure that is not merged yet within the structure.

In the end, we would briefly consider a “processing-based” perspective, from which an island violation is simply a matter of performance. Klunder (2004), among many others, suggests that the unacceptability of subextraction (also) from subjects depends mainly on the special status of

---

15 As discussed in Jurka (2010:151), the PP in (8) might be base-generated in a hanging topic position (paraphrasing (8).b: “of which car was it the case that the driver of this car was awarded a prize?”); it goes without saying that this solution would not explain the contrast between (4).a and (4).b since also (4).a a paraphrase would lead to a grammatical sentence: “of which car was it the case that the driver of this car caused a scandal?”.
16 Following Longobardi (1991), the presence of a definite article could suggest that the wh- phrase is not really extracted but directly generated in a topic-like position leading to a ‘possessive’ interpretation within the subject (where the possessor is implicit). This solution however does not explain the contrast discussed in Bianchi & Chesi (2014): partitive subjects permit the resumptive “possessor strategy” (iii).a but not for extraction (iii).b:
   (iii) a. A proposito di Maria, ci ha telefonato uno dei figli
   speaking of Maria us has phoned one of the sons (Speaking of Maria, one of her sons called us)
   b. *Maria, di cui ci ha telefonato uno dei figli
   Maria of whom us has phoned one of the sons (Mary, one of whose sons called us)
17 Following Cinque (1990), we should expect a DP, not PP, to be able to exploit the proposed null resumptive pronoun strategy in case like (4).b, that according to Cinque (1990) are just apparent cases of extraction.
18 As for the base vs. derived subject position, the example (4).b suggests that also the subject in a derived position can permit subextraction; as for the internal vs. external argument, the relevant counter example is provided by Jiménez Fernández (2009):
   (iv) De qué cantante te parece que [algunas fotos _ ] han escandalizado a la audiencia
   of which singer you seem that some pictures have shocked to the audience
the pre-verbal subject in SV languages: first, as attested by the preference for subject over object wh-dependency formation (see §3.1), the subject constituent introduces a “memory load” whenever a non-local dependency spanning over it must be established; second, the subject also induces an “integration cost” in terms of new referents to be computed (Gibson 1998, §3.1). Despite the fact that these considerations cannot help, by themselves, in accounting for the contrasts in (8) (using Kluender’s terminology, both “memory load” and “integration cost” are constant in (8)), this suggests (again) that RS performance can be explained in terms of “memory usage” looking at phrase structure “from-left-to-right”.

4 A matter of degrees: reconciling competence and performance

While distinguishing between grammaticality and acceptability Chomsky (1965:11) observes that both concepts are gradable, in the sense they can be assessed on scales. Even though grammaticality and acceptability scales do not coincide, being grammaticality just one of the many aspects contributing to acceptability (the other being semantic plausibility, pragmatic adequacy and intonation for instance), it is purely an experimental exercise to keep such external factors constant (for instance removing semantic, pragmatic and intonational biases) while looking at subtle structural variables correlations, if any, with respect to RS performance. This is how recently fine grained distinctions have been attested using grammaticality judgments (Jurka 2010, Bianchi and Chesi 2014 a.o.). This permitted to consider seriously systematic evidence where, in the past, we simply found performance-related noise.

The emerging picture is, from many perspectives, even simpler and psycholinguistically more reasonable. For instance, assuming that our algorithmic derivation proceeds Top-Down (Chesi 2014) both self-embedding and subject island “violations” can be accommodated. On one side, we obtain a left-right (that is real-time based) derivation (this is because external merge is regulated by selection features on the already processed lexical items); this fulfills the “processing-based” assumption for which the computation of constituents must be ordered roughly from left to right (§3.1, Chesi 2014). On the other side, we can easily decide if a certain selection requirement introduced by a constituent (e.g. the possessor requirement associated, in generation, to a given noun phrase within the subject pre-verbal position) must be readily processed (this is marked by the presence of the stranded preposition) or must wait to be interpreted within the predicative nucleus (i.e. “reconstruction” takes place, hence deriving extractability and the problematic preference for pied-piping, Bianchi and Chesi in press).

Notice that also non-feature based movement can be reconciled with a top-down movement perspective: according to Moro (2000) movement is not triggered by any specific feature (see Chesi 2014 for the critical problems a feature-based movement approach causes from an algorithmic perspective) but by a “symmetric” geometrical configuration that does not permit to determine univocally the relevant (asymmetric) C-command relations between elements. From a reverse (top-down) perspective, we expect that symmetric structures must be recreated during the course of the derivation (e.g. to assign thematic roles) and that certain “asymmetric” configurations should trigger movement. We interpret “asymmetry” as the mismatch between the expected (i.e selected) features and the actual features introduced in the derivation (following Chesi 2014, unexpected features force movement) and we assume that a “symmetric” configuration is created, when the pending expectation (i.e. selection) is fulfilled (that is, when the relevant holding constituent is “reconstructed” in an appropriate thematic position).
We conclude by mentioning that the idea behind a Top-Down approach is not new at all: Yngve (1960) suggested a similar derivational constraint to explain, from a competence perspective, cases like center-embedding. Chomsky (1965:197) liquidates this approach by saying that he sees “no plausibility at all to the assumption that the speaker must uniformly select sentence type, then determine subcategories, etc. finally, at the last stage, deciding what he is going to talk about; or that the hearer should invariably make all higher-level decisions before doing any lower-level analysis”. Here we do not want to compare our approach with Yngve’s one (the two are clearly different in many respects), but we want to stress that any theory of grammar that is mechanical and derives the correct (graded) oppositions without any uncertainty is totally plausible and cannot be excluded a priori. In this sense, we see no difference in a top-down or bottom-to-top approach since they are both “competence-based” assumptions.

References


A Source of Parametric Variation in the Lexicon

Guglielmo Cinque
Università Ca’ Foscari, Venezia

1 Introduction

An influential conjecture concerning parameters is that they can possibly be “restricted to formal features of functional categories” (Chomsky 1995, 6; cf. Borer 1984, 2f.). In Rizzi (2009, 2011) such features are understood as instructions triggering one of the following syntactic actions:

(1) a. External Merge
    b. Internal Merge (Move)
    c. Pronunciation/non-pronunciation (the latter arguably dependent on Internal Merge – Kayne 2005a)

Here I discuss what appears to be a particularly pervasive source of variation among languages in the domain of the lexicon (both functional and substantive) and consider whether and how it can be reduced to one of the above actions.

The variation can be preliminarily characterized as follows: language A has two (or more) lexical items which correspond to just one lexical item in language B.

2 Functional Lexicon

Example 1 (Zanuttini 1997, §3.3.1 and §3.3.2)
The Piedmontese northern Italian dialect of Turin has two sentential negative markers: nen, which is a neutral negative marker (it simply negates a certain proposition P), and pa,

---

1 I wish to thank for their comments to an oral presentation of this work Richard Kayne, Iliyana Krapova, Rita Manzini, Henk van Riemsdijk, Luigi Rizzi, and Peter Svenonius.
corresponding to standard Italian *mica*, which is a presuppositional negative marker (it negates a certain proposition \( P \) that the speaker believes to be presupposed in the context of utterance, with the effect of denying the correctness of such a presupposition). \(^2\)

The northern Italian Valdotain dialect of Cogne, on the other hand, has only one sentential negative marker, *pa*, which covers both functions (it can be used either as a neutral or as a presuppositional negative marker).

**Example 2 (Cinque 1999, §4.19, §4.25 and 208 fn. 57)**
The two English adverbs *soon* and *early* seen in (2)a. and b. are rendered in Italian by the single adverb *presto*, as seen in (3)a.-b.: \(^3\)

(2) a. He will **soon** have to get up  
   b. He will have to get up **early**  

(3) a. **Presto** si dovrà alzare  
   ‘He will **soon** have to get up’  
   b. Si dovrà alzare **presto**  
   ‘He will have to get up **early**’

### 3 Substantive Lexicon

**Example 1:**
While Italian has separate lexical items to refer to ‘arm’ and ‘hand’, *braccio* and *mano* respectively, and ‘leg’ and ‘foot’, *gamba* and *piede* respectively, Bulgarian uses one lexical item for both ‘arm’ and ‘hand’, *raka*, \(^4\) and one lexical item for both ‘leg’ and ‘foot’, *krak*.

**Example 2**
To the distinct English lexical items *grandson/granddaughter* (i.e. male/female grandchild) and *nephew/niece* (i.e. male/female child of sibling), only one lexical item corresponds in Italian: *nipote*, for ‘grandson’/’granddaughter’/’nephew’/’niece’.

The examples could easily be multiplied.

### 4 The Logic Underlying this Pattern of Variation

I take this pattern of variation not to be accidental, and to arise from the fact that the functional or substantive denotata of the two (or more) lexical items of language \( A \) which correspond to the

---

\(^2\) For a discussion of the pragmatic conditions required for such presuppositional negation to be felicitous see Cinque (1976) and Zanuttini (1997, Chapter 3).

\(^3\) I take these adverbs (more exactly, adverbial phrases) to belong to the functional lexicon as they appear to correspond in terms of position and interpretation to two independent aspectual projections (cf. Cinque 1999, Chapter 3).

\(^4\) For the lexicalization of ‘hand’ and ‘arm’ cross-linguistically the World Atlas of Language Structures (http://wals.info/chapter/129) reports that 228 languages have an identical word and 389 languages have two different words.
unique lexical item of language B share one component/feature (while differing with respect to other components/features). Language B capitalizes on this shared component/feature. That is, the single lexical item of language B is uniquely specified for the common component/feature and left unspecified (in ways to which I return) for the differentiating components/features. Language A, on the other hand, capitalizes on the differentiating components/features. That is to say, its two (or more) lexical items corresponding to the single lexical item of language B are specified both for the shared component/feature and for the differentiating component(s)/feature(s).

Consider in this light the examples of the functional lexicon given in section 2 (§4.1) and those of the substantive lexicon given in section 3 (§4.2).

### 4.1 The Functional Lexicon

Concerning the first example of section 2, we noted there, after Zanuttini (1997, §3.3.1 and §3.3.2), that the Piedmontese of Turin has two sentential negative markers, the presuppositional *pa* and the neutral *nen*. The two, in addition to the different pragmatic conditions that govern them, also differ in the position they occupy within the clause. The presuppositional negative marker *pa* necessarily precedes an adverb like *gia* ‘already’, while the neutral negative marker *nen* necessarily follows it:

(4) a. A l’è *pa/*nen gia parti
   Cl Cl is neg already left
   ‘He hasn’t already left’

b. A l’avia *gia nen/*pa salutami cul di la
   CL CL had already neg greeted-me that day there
   ‘Already on that day he had not greeted me’

The overall order given in Zanuttini (1997, 72) is thus *pa* > *gia* > *nen* (> *sempre*). In fact, the two can co-occur, in the expected order (*pa nen*), as shown in (5):

(5) Fa *pa nen* (*nen pa*) sulì
   do neg neg that
   ‘Don’t do that!’ (the assumption is that the addressee is about to do it)

The Valdotain dialect of Cogne, on the other hand, has only one negative marker, which can be used either as a presuppositional or as a neutral negation: *pa*. However, this is not merely a lexical quirk. When it is presuppositional *pa* precedes *dza* ‘already’; when it is neutral it follows *dza*:

(6) a. L’è *pa dza* parti?
   Cl Cl is neg already left
   ‘He hasn’t already left, has he?’

b. I m’a *dza pa* saluià ce dzor lai
   CL me has already neg greeted that day there
   ‘Already that day he didn’t greet me’
The overall order given in Zanuttini (1997,82) is thus \( pa > dza > pa \) (> toujou).\(^5\)

All of this seems to me to point to the presence of two specialized negative positions (or projections), which share a common core (Negation of P); one below the projection occupied by the adverb ‘already’, expressing simple Negation of P, and one above it, expressing Negation of P, where P is presupposed (denial of P).

If the lexical specifications of Piedmontese *nen* and *pa* are \{Negation of P\} and \{Negation of P, P presupposed\}, respectively, each will be uniquely matched with the corresponding projection. If on the other hand the lexical specification of Valdotain *pa* is \{Negation of P\}, with unspecified \{P presupposed\}, then it will be able to match either projection.\(^6\)

\[\begin{align*}
\text{(7) a. Syntax:} & \quad F_1 \quad > \quad gia \text{‘already’} \quad > \quad F_2 \\
& \quad +\text{Neg P} \\
& \quad +\text{P presupposed} \\
\text{b. Lexicon:} & \quad \begin{array}{ll}
\text{Piedmontese} & \text{Valdotain} \\
\text{pa:} & \text{pa:} \\
& +\text{Neg P} \\
& +\text{P presupposed} \\
& +/-\text{P presupposed} \\
\text{nen:} & +\text{Neg P}
\end{array}
\end{align*}\]

Consider now the second example of section 2, concerning the Italian adverb *presto*, which corresponds to both English *soon* and *early*. The relevant examples, (1) and (2), are repeated here as (8) and (9):

\[\begin{align*}
\text{(8) a. He will } & \text{soon have to get up} \\
\text{b. He will have to get up } & \text{early} \\
\text{(9) a. } & \text{Presto si dovrà alzare} \\
& \text{‘He will soon have to get up} \\
& \text{b. Si dovrà alzare } \text{presto} \\
& \text{‘He will have to get up early’}
\end{align*}\]

When *presto* precedes the finite verb it is interpreted as ‘soon’. When it follows the verb it is interpreted as ‘early’.\(^7\) The two *presto* can co-occur:

---

\(^5\) It is not clear whether the two *pa* can co-occur (Raffaella Zanuttini, p.c.).

\(^6\) This requires not extending to such cases of underspecification the Aspects proposal that “each lexical entry automatically, by convention, contains the feature [-A] for every lexical category A, unless it is explicitly provided by the feature [+A].” (Chomsky 1965,111). The notion of ‘underspecification’ of syntactic features discussed here is different from the phonological and (one type of) morphological notion of ‘underspecification’ discussed in the literature; namely, that concerning those features (like the aspiration of onset stop consonants in English) that are predictable and thus can be expunged from the lexical representation and added through a rule (cf., among others, Archangeli 1984, Farkas 1990, and Steriade 1995). In the cases discussed here the underspecified features are crucially not added (specified) at all, whether by rule or otherwise. It does bear some similarity however with the notion of underspecification employed in Distributed Morphology to account for cases of syncretism. According to Halle’s (1997) Subset Principle “the phonological exponent of a Vocabulary Item is inserted into a position if the item matches all or a subset of the features specified in that position.” (also see Embick and Noyer 2007,§2.4).

\(^7\) In English, *early* also has to follow the verb:

(i) He <*> early > got up <*> early>
(10) **Presto** si dovrà alzare **presto** ‘He will soon have to get up early’

In Cinque (1999) I had suggested that **presto**, qua ‘soon’, encodes “the fact that an event is going to take place a short while after some reference time” (p.97) (cf. **Si rese**/renderà conto **presto** che lo stavano/stanno imbrogliando ‘He soon realized/will soon realize that they were/are cheating him’). **Presto**, qua ‘early’, appears instead to be paraphrasable as ‘a short time after the beginning of a scale of waking-up (more generally: V-ing) times’ (cf. **Si è alzato presto** ‘He woke up early’). The shared core-component/feature of the two functional projections thus appears to be a short time after x. If the lexical specification of **presto** is [a short time after x] (x left unspecified), then **presto** will be able to match the two distinct functional projections: the one associated with [in a short time after x, x a reference time] and the other associated with [in a short time after x, x the beginning of a scale of V-ing times].

Consider next the examples from the substantive lexicon mentioned in section 3.

### 4.2 Substantive Lexicon

The items of the substantive lexicon have components/features that, differently from those of the functional lexicon, do not match components/features of functional heads. Their components/features rather appear to match the categories with which we interpret/represent the world, broadly taken.

Consider the Italian – Bulgarian contrast shown in example 1 of section 3. While Italian has two separate lexical items for ‘arm’ and ‘hand’ (**braccio** and **mano**, respectively), Bulgarian has a single lexical item, **raka**, which can refer to either ‘arm’ or ‘hand’. Similarly, while Italian has two separate lexical items for ‘leg’ and ‘foot’ (**gamba** and **piede**, respectively), Bulgarian has just one lexical item, **krak**, which can refer to either ‘leg’ or ‘foot’. I take this to suggest that Bulgarian expresses just the shared component/feature of ‘arm’ and ‘hand’ (namely, ‘upper limb’), and ‘leg’ and ‘foot’ (namely, ‘lower limb’), leaving unspecified what further differentiates ‘arm’ from ‘hand’ and ‘leg’ from ‘foot’. The separate lexical items of Italian for ‘arm’ and ‘hand’ and ‘leg’ and ‘foot’, on the other hand, in addition to specifying the shared component/feature, also specify what differentiates ‘arm’ from ‘hand’ and ‘leg’ from ‘foot’. The lexical specifications of the different lexical items of the two languages can thus be represented in first approximation as in (11):

\[
\begin{align*}
\text{a. Italian:} & \quad \text{braccio ‘arm’ (+upper limb, - extremity)} \\
& \quad \text{mano ‘hand’ (+upper limb, - (+upper limb, - extremity))} \\
& \quad \text{gamba ‘leg’ (+lower limb, - extremity)} \\
& \quad \text{piede ‘foot’ (+lower limb, - (+lower limb, - extremity))} \\
\text{b. Bulgarian:} & \quad \text{raka ‘arm’ or ‘hand’ (+upper limb)} \\
& \quad \text{krak ‘leg’ or ‘foot’ (+lower limb)}
\end{align*}
\]

Consider now the second example of section 3. In Italian, a single lexical item, **nipote**, corresponds to English **grandson, granddaughter, nephew and niece**; abstracting away from the
male/female distinction (also present in Italian in the determiners that precede the noun: *un/il* (masc.) *nipote*, *una/la* (fem.) *nipote*, *nipote* apparently corresponds in English to two distinct kinship relations, which can be represented as in (12):

**English:**

(12) a. anchor/ego  
   |  
   | (son/daughter)  
   
   |  
   | grandson/granddaughter

b. anchor/ego ——— (brother/sister)  
   |  
   | nephew/niece

1st line: + descending, - ascending  
2nd line: + descending, - ascending  

1st line: - descending, - ascending (= horizontal)  
2nd line: + descending, - ascending

These two kinship relations have, nonetheless, something in common. A degree 2 distance from the anchor/ego. The relation can be made identical if one suspends the directionality of the first line. By leaving unspecified its “descending” component/feature, one can collapse the two kinship relations into one, as in (13), which is precisely what Italian seems to do.9

**Italian:** *nipote*

(13)  

1st line: +/- descending, -ascending  
2nd line: + descending, - ascending

5 Underspecification vs. Silent Elements

In taking an ‘underspecification’ parametric approach to cross-linguistic differences in the lexicons of languages care should be taken to distinguish cases amenable to it from cases arguably involving the presence of silent elements (in one language but not in another), as in Richard Kayne’s recent work. Consider another difference between Italian and Bulgarian, which at first sight appears to be of the same ilk of the preceding ones. While Italian has one word,

---

9 If one takes the +descending value of the first line, one gets the ‘grandson/granddaughter’ meaning; if one takes the -descending value, one takes the ‘nephew/niece’ meaning.

It is tempting to take such under-specification of components/features as a way of capturing the cross-linguistic typology of kinship terms. To mention just one example, in Western Dani (Papuan, Trans-New Guinea - Barclay 2008,61), the lexical word *ombo* means both ‘grandparent’ and ‘grandchild’. In English *grandparent* and *grandchild* have two degrees of distance from the anchor/ego. In the former, both lines are +ascending -descending; in the latter both are -ascending +descending. Western Dani *ombo* thus appears characterizable as underspecified for the +/-ascending, +/-descending components/features (provided that both lines have the same value for such components/features). This line of analysis makes us expect that no single term may cover, say, ‘grandchild’ and ‘cousin’, or ‘nephew/niece’ and ‘cousin’, or ‘grandchild’, ‘nephew/niece’ and ‘cousin’ (‘cousin’ being 3 degrees of distance: 1) + ascending, - descending; 2) - ascending, - descending; 3) - ascending, + descending).
**molto**, for ‘(very) much’ and another word, *troppo*, for ‘too much’, Bulgarian has a single word, *mnogo*, for both. See, for example the contrast between (14)/(15) and (16):

**Italian:**
(14)  
a. Non ho bevuto **molto**  
‘I didn’t drink much’  
b. Ha **molti** libri  
‘(S)he has got many books’  
c. Suo figlio è **molto** stupido  
‘His son is very (*much) stupid’

(15)  
a. Ho bevuto **troppo**  
‘I drank too much’  
b. Ha **troppi** libri  
‘(S)he has too many books’  
c. Suo figlio è **troppo/molto** stupido per fare una cosa del genere  
‘His son is too/very stupid to do such a thing’

**Bulgarian:**
(16)  
a. Toj pie **mnogo**  
‘He drinks very much or too much’  
b. Toj ima **mnogo** knigi  
‘He has many or too many books’  
c. Sinăt mu e **mnogo** glupav  
‘His son is very or too stupid’

c’. Sinăt mu e **mnogo** glupav za da razbere tova  
‘His son is too stupid to understand that’  
(another option is to use **tvärde** ‘too (much)’: Sinăt mu e **tvärde** glupav za da razbere tova)

In this case, thinking of Kayne (2005b, §3.5, 2007), there is reason to believe that the apparent ambiguity of *mnogo* ‘very (much/many)/too (much/many)’ is due to the presence of either one of two different silent degree words (*strašno* ‘very’ and **tvärde** ‘too’), as these are the only degree words which are optional in the paradigms (17) and (18), and the only two which are in complementary distribution with *mnogo* in the paradigm in (19). Consider the following paradigms (Iliyana Krapova, p.c.):

(17)  
a. *(kolko)* mnogo ‘how much’ (or simply **kolko**)  
b. *(tolkova)* mnogo ‘so much’  
c. *(pò)* mnogo ‘more’ (or **poveče**)  
d. *(naj)* mnogo ‘most’  
e. *(strašno)* mnogo ‘very much/many’  
f. *(tvärde)* mnogo ‘too much/many’

(18)  
a. *(kolko)* mnogo knigi ‘how many books’ (or simply **kolko knigi**)  
b. *(tolkova)* mnogo knigi ‘so many books’  
c. *(pò)* mnogo knigi ‘more books’
Thus the ambiguity of (16)a–c is plausibly to be attributed to the presence of a silent degree word; either strašno ‘very’ or tvârde ‘too’ (which cannot be overtly realized within an AP, if mnogo is). Here mnogo is not lexically underspecified. It acquires its apparent ambiguity as a consequence of the independent property of strašno ‘very’ and tvârde ‘too’ to be unpronounced.

6 Conclusions

Returning now to the question posed at the beginning (whether and how the cases that we have examined so far can be reduced to one of the parametric actions seen in (1) above), it appears that while the contrast between Italian molto/troppo vs. Bulgarian mnogo is indeed amenable to the action in (1)c (pronunciation vs. non-pronunciation), the other cases examined in sections 2, 3, and 4 must be attributed to an additional parametric action: underspecification of syntactic features in the (substantive and functional) lexicon.

References


10 I.e., strašno MNOGO glupav ‘very stupid’.
11 I.e., STRAŠNO mnogo glupav ‘very stupid’.
12 I.e., tvârde MNOGO glupav ‘too stupid’.
13 I.e., TVÂRDE mnogo glupav ‘too stupid’.


1 Introduction

The goal of this paper is to investigate the ambiguity of sentences like (1):

(1) At the party, I saw three boys who I know and one girl.

On one interpretation, I propose (1) has the following structure:

(2) At the party, I saw three boys who I know and one girl <who I know>

The angled brackets <…> indicate that the enclosed string is present syntactically but not pronounced. I will call the process by which the relative clause in (2) is not pronounced relative clause deletion.

The interpretation of (1) with the structure in (2) can be brought out by the following situation (situation A). Suppose you go to a party, and there are twenty people: ten boys and ten girls. Furthermore, you see all twenty people. Suppose out of the twenty people, you know three boys and one girl. So in particular, you see ten girls, but you only know one of them. Then (1) with structure (2) is true in that situation.

Sentence (1) also has another structure where one girl is not modified by a covert relative clause. Consider situation B. Suppose you go to a party with 11 people: ten boys and one girl. Suppose that you see all eleven people. You happen to know three of the boys. Furthermore, you do not know the girl. Then (1) is also true in this situation. But in this case, the structure is different from (2). Rather, when (1) is true in situation B, there is no covert relative clause modifying one girl.

Crucially, (1) with structure (2) is false in situation B (since you do not know the girl in situation B). And (1) with no covert relative is false in situation A (since you see more than one girl). So this makes it easy to distinguish the two structures of (1) truth conditionally.
2 Comparatives

The example in (1) involves a coordinate structure. Another common source of deleted relatives is comparatives.

(3) At the party, there were more girls who I know than boys.

This sentence is ambiguous in much the same way as (1). On one interpretation it is equivalent to:

(4) At the party, there were more girls who I know than there were boys who I know.

On another interpretation, (3) is equivalent to:

(5) At the party, there were more girls who I know than there were boys (known or unknown).

Such comparatives with deleted relatives are easy to find on the internet:

(6) In eastern European countries, there are more boys who smoke than girls.  
(https://www.uv.es/lisis/sofia/buelgahandbook.pdf)

(7) Cats are high on the kill list because there are many more cats who need homes than dogs.  

(8) There are as many boys who play with the dolls as girls.  
(http://fahnlanders.blogspot.com/)

(9) “Even in Europe, there are many fewer boys who study ballet than girls,” Alberto observes.  
(http://balletnebraska.org/2012/a-dancers-journey-alberto-liberatoscioli/)

In another kind of comparative structure, the head of the relative clause is not the compared element. In this case as well it is possible to find internet examples (all of which have the relevant interpretation for me):

(10) Girls who've experienced parental separation are more likely than boys to turn to the Non-Productive coping strategy of self-blame to deal with uncontrollable family stressors.  

(11) …and that girls who dropout are more likely than boys to return.  
(http://books.google.com/books?id=esksAAAAIAAJ)
(12) My figures suggested that girls who are dyslexic or dyspraxic are much more likely than boys to escape being diagnosed at school.

(13) Men who are widowed are more likely than women to remarry.

(14) Children who are exposed to iodine-131 are more likely than adults to get cancer later in life.
(http://newamericamedia.org/2011/03/are-there-health-risks-in-us-from-japans-nuclear-disaster.php)

(15) In fact, according to the Environmental Protection Agency, a child who is exposed to a carcinogen is 10 times more likely to develop cancer than an adult.
(http://www.babble.com/mom/home-sweet-home-a-toxic-world-for-baby/)

I assume that the analysis of (13) involves relative clause deletion as illustrated in (16):

(16) Men who are widowed are more likely to remarry than women <who are widowed> <are likely to remarry>.

3 Identity and Parallelism

In all cases so far, the deleted relative clause is syntactically identical to its antecedent (see van Craenenbroeck, Jeroen and Jason Merchant 2013 for a recent discussion of the issue of syntactic vs. semantic identity conditions on ellipsis).

(17) Syntactic Identity
A relative clause R is deleted under syntactic identity with an antecedent relative clause.

Syntactic identity accounts for the following contrast:

(18) a. There are more women who are married than men.
   b. There are more women who have a husband than men.

Given relative clause deletion, (18a) can mean that there are more married women than married men. However, (18b) does not have the interpretation “than men who have a wife”, even though that is a plausible DP to compare with “women who have a husband”. Such an interpretation for (18b) is ruled out by the identity condition in (17), since the relative clause “who have a husband” is not identical to “who have a wife”. Rather, (18b) either means (i) that there are more women who have a husband than men all together (no relative clause deletion), or (ii) that there are more women who have a husband than men who have a husband (relative clause deletion).
In addition to (17), there also seems to be a parallelism condition to the effect that the deleted relative clause and its antecedent must be found in the same syntactic environment:

(19) Parallelism
Relative clause deletion can only take place in the following structure:

\[ [\text{XP}_1 \ldots [\text{Head}_1 \ \text{Antecedent}] \ldots ] \ \text{and/or} \ [\text{XP}_2 \ldots [\text{Head}_2 <\text{Relative Clause}>] \ldots ] \]

where Head2 is focused and XP1 is a member of F(XP2), the focus value of XP2.

The two XPs in (19) differ in the NP heads of their respective relative clauses (including modifiers): Head1 and Head2, where Head2 (and perhaps also Head1) is focused. I understand F(XP), the focus value of XP, as in Fox 1999 (see section 7 below).

Given (19), one issue to investigate is whether Head1 and/or Head2 are marked prosodically when there is relative clause deletion. In other words, in a sentence like (1) are the two interpretations distinguished prosodically? Is one girl stressed in a particular way when the relative clause is deleted?

When the parallelism condition is not met, relative clause deletion is not generally possible with indefinite determiners:

(20) One boy who I know was talking to three girls.
(no relative clause deletion interpretation)

(21) Some boy who is in my class was talking to three girls on the steps.
(no relative clause deletion interpretation)

(22) Boys who drop out usually blame it on girls.
(no relative clause deletion interpretation).

But much more work is needed to nail down the specific conditions on when relative clause deletion is possible.

4 Deletion of other Modifiers

While the focus of this paper is on relative clause deletion, it is clear that other modifiers can be deleted in the same way. Consider the following example:

(23) Some boy from every school reads comic books, and some girl does too.

The first conjunct in (23) can have an inverse linking interpretation (see May 1977) where for each school, there is some boy who reads comic books. Furthermore, if the first conjunct has an inverse linking interpretation, then so does the second conjunct. The presence of an inverse linking interpretation for the second conjunct suggests the following underlying representation:

(24) …and some girl <from every school> does <read comic books> too.
Relative Clause Deletion

I have not systematically investigated deletion of prenominal adjectives, but preliminary data shows that it may sometimes be possible:

(25)  a. There were more tall girls than boys in my class.
     b. Tall girls are more likely than boys to play basketball.
     c. I saw three tall girls and one boy.

(26)  a. I taught more foreign girls than boys.
     b. Foreign men are more likely to remarry than women.
     c. At the party, I saw one foreign boy and two girls.

The modifier deletion interpretation is possible in the (a) sentences, difficult in the (b) sentences and impossible in the (c) sentences. I have no account of this variation.

5 Definite Determiners

In (1) the relevant DPs were both indefinites. Relative clause deletion is also possible with definite DPs in examples like the following:

(27)  a. I saw the boy who I know, and the girl too.
     b. I saw the one boy who I know, and the three girls too.
     c. I saw the boy who I met in Paris, and the girl too.
     d. I saw the boy who I met in Paris, but not the girl.

(27d) is derived from (28) by deleting everything following but other than the negation and the girl:

(28) I saw the boy who I met in Paris but I didn’t see the girl who I met in Paris.

Relative clause deletion is also possible with definite DPs in comparatives. The following example can have the interpretation “more likely to remarry than women who were widowed.”

(29) The men who were widowed were more likely to remarry than the women.

In some cases involving a definite determiner, it is unclear whether there is relative clause deletion, or a semantic process of domain restriction (on the issue of domain restriction with definite DPs see Heim 2011). Consider (30):

(30) I met a boy and a girl in Paris. The girl was a teacher.

In this sentence, the DP the girl in (30) seems to have a contextually specified domain narrowing of the set of all girls (the denotation of girl) to the set containing just the girl who I met in Paris. One possibility is that this domain narrowing is a semantic process distinct from the syntactic process of relative clause deletion. Domain narrowing would be purely semantic, involving no deleted relative clause. If the domain narrowing approach to (30) were correct, we
would have to question whether relative clause deletion were involved at all in any of (27) and (29), and perhaps even in the cases with an indefinite determiner.

Another possibility is that (30) also involves relative clause deletion. So that the underlying structure of (30) is really:

(31) I met a boy and a girl in Paris. The girl <who I met in Paris> was a teacher.

On the relative clause deletion analysis of (31), the deleted relative clause does not have a syntactically identical antecedent, since there is no previous relative clause of the form [who I met in Paris]. For the same reason, the deletion in (31) would violate the parallelism condition (19).

A related fact is that relative clause deletion readings are much easier with definite determiners in sentences like (20-22) above:

(32) The one boy who I know was talking to the three girls.

This sentence easily admits the following interpretation:

(33) The one boy who I know was talking to the three girls who I know.

Faced with these data, one option would be to conclude that there is a semantic process of domain narrowing (distinct from the syntactic process of relative clause deletion) found with definite DPs. Alternatively, it may be the case that the conditions in (17) and (19) do not hold for relative clause deletion with definite DPs. While I am unable to resolve this issue here, the following section gives independent syntactic evidence that relative clause deletion is possible in certain cases with definite DPs.

6 Reconstruction

In this section, I provide syntactic evidence for relative clause deletion. In the following example, it appears that the reflexive in picture of himself in the woods is bound by John.

(34) The picture of himself in the woods that John took is prettier than the picture of himself on the boat.

We can understand this in terms of the copy theory of movement, as illustrated in (35a). I assume that OP is the covert equivalent of the relative pronoun which. Furthermore, as demonstrated in section 9, the CP complement of the is dominated by another projection RelP (into which the NP picture of himself raises), which I leave out here for simplicity. As shown in (35b), the restriction of the higher copy is deleted, and the OP of the lower copy is deleted and replaced by a variable (see Chomsky 1995: 202-212). Fox 2003 claims that there is syntactic rule (“Trace Conversion”) which converts the lower copy into a definite description. I will not pursue this issue here.

(35) a. the [CP [OP picture of himself] that John took [OP picture of himself]]
b. the \([\text{CP} \text{OP}_x \text{that } \text{John}_1 \text{took } [x \text{picture of himself}]])

In the LF representation of (35b), the lower copy contains the reflexive, which is bound by the subject. But then by the same reasoning, the NP *picture of himself on the boat* in (34) must also be modified by a covert relative clause allowing for reconstruction.

A similar example can be constructed based on bound variable pronouns:

(36) I saw the picture of his father that every boy took,
    but not the picture of his mother.

In (36) the universal quantifier is the subject of a finite relative clause. Furthermore, the pronoun in *picture of his father* is interpreted as a variable bound by *every boy*. Lastly, there is an entailment that I saw several pictures, one for each boy. Such an entailment suggests that *every boy* undergoes QR out of the finite relative clause (see Huley and Sauerland 2006: 132). But QR of *every boy* over the head of the relative clause *picture of his father* would yield a WCO violation. Because of this, Huley and Sauerland 2005: 134 propose that such examples involve reconstruction of the head of the relative clause. The resulting structure can be represented as in (37):

(37) \([(\text{every boy}) \text{y} [\text{I saw the } [\text{CP} \text{OP}_x \text{that } \text{y took } [x \text{picture of y’s father}]]])]

But then, given that the pronoun in *picture of his mother* in (36) is interpreted as a bound variable, by parity of reasoning *picture of his mother* should also undergo reconstruction into a relative clause (out of which QR has taken place). Such a relative clause is made available by relative clause deletion: *the picture of his mother* \(<\text{that every boy took}>.*

### 7 Strict and Sloppy Readings

An example of the distinction between strict and sloppy interpretations of a pronoun in an elided VP is given in (38) (see Ross 1986[1967]: 207 who first noted the distinction):

(38) John loves his mother and Bill does too.
    Sloppy: Bill loves Bill’s mother.
    Strict: Bill loves John’s mother.

In the case of relative clause deletion, the following examples illustrate sloppy interpretations:

(39) a. I met three freshmen who liked their professors and two sophomores.
    b. I met more freshmen who like their professors than sophomores.
    c. Freshmen who like their professors are more likely to attend class than sophomores.
All three sentences in (39) admit relative clause deletion with a sloppy interpretation of the deleted pronoun. For example, the relevant set of sophomores in (39b) are those who like their own professors.

Other examples show strict interpretations:

(40) I like the picture of John that was on his mother’s fridge, but not the picture of Bill.

The most natural interpretation of (40) is the following:

(41) I like the picture of John that was on John’s mother’s fridge, but
I do not like the picture of Bill that was on John’s mother’s fridge.

Recall the identity condition on relative clause deletion in (17). With the exception of the data in section 5 involving definites, such a condition handles all the data given in the paper so far. For example, (39b) will have the following analysis:

(42) I met more freshmen who₁ like their₁ professors than sophomores <who₁ like their₁ professors>.

Since the second relative clause is syntactically identical to the first, it can be deleted. The indices in (42) are not interpreted in terms of coreference, but rather variable binding (who₁ binds their₁ syntactically and there₁ is interpreted as a bound variable). So there is no implication that their₁ in the first relative clause refers to the same person as their₁ in the second relative clause. See Sag 1977[1976]: 74, 86-102 for a related treatment of the identity condition on VP deletion.

Consider now an example where the antecedent of the pronoun is not contained within the deleted relative clause. In this example, John and Bill are wildlife biologists who tag wild animals and do follow up investigations of the animals they tag. I report on their activities by saying:

(43) John spotted one lion that he tagged last week, and Bill two zebras.

On one interpretation, this can be true if both John and Bill spotted many lions and zebras, but only a few that they actually tagged. On that reading, there is relative clause deletion, represented as follows (the verb spotted is deleted by gapping):

(44) John₁ spotted one lion that he₁ tagged last week,
and Bill₂ <spotted> two zebras <that he₂ tagged last week>.

Crucially the index of the pronoun in the deleted relative is different from the index of pronoun in the non-deleted relative (since their antecedents John₁ and Bill₂ have different indices). The question then is whether (44) violates the syntactic identity condition in (17). There are a number of possible ways to deal with sentences like (44) that have been proposed in the literature. First, Fiengo and May (1994: 95) propose the syntactic notion of i-copy, where the dependency in the first clause is identical to the dependency in the second clause, even though the indices are different. Second, Rooth (1992) proposes to simply ignore indices for the purposes of the syntactic relation of “reconstruction” (“pronominal indices may vary”), constraining the interpretation of the deleted VP by a semantic condition on focus. Third, Fox
(1999) proposes to eliminate the syntactic identity condition all together, and to force syntactic identity through a parallelism condition on focus. On all of three of these approaches, the non-identity of the indices in (44) would not block deletion.

Consider Fox (1999) in more detail. According to Fox’s parallelism condition, the antecedent (S1 = [John1 spotted one lion that he1 tagged last week]) has to be in the focus value of the second clause S2. The focus value of S2 is defined as the set of sentences which are alternatives to S2: \( F(S2) = \{ S : \exists x \exists y [ S = x \text{ spotted } y \text{ that } x \text{ tagged last week} ] \} \), where x ranges over DPs that are alternatives to Bill and y ranges over NPs that are alternatives to two zebras. S1 is an element of F(S2), since S1 is of the syntactic form [x spotted y that x tagged last week] (assuming John1 and he1 both count as x). Therefore, parallelism is satisfied. A consequence of this approach is that indices of pronouns which are not bound by focused DPs have to match exactly (since if there were mismatched indices, S1 would not be in F(S2)). In other words, both syntactic identity, and the exceptions to it (as in the mismatched indices of (44)) follow from Fox’s parallelism condition.

As noted, Fox argues against a syntactic identity condition, subsuming it under the parallelism condition on focus. However, syntactic identity is implicit in a number of places in his system (e.g., in the definition focus values, see footnotes 3 and 4 and the parallelism condition). But the question of whether or not deleted structures are syntactically identical to their antecedents is non-trivial (see van Craenenbroek and Merchant 2013 for an extensive survey of the issues). And different answers to the syntactic/semantic identity question could lead to different formulations of the parallelism condition, and vice versa. Therefore, I will continue to assume that there is a distinct syntactic identity condition (17), and that it should be interpreted as allowing mismatched pronominal indices only when the relevant pronouns are bound by focused DPs (allowing (44)).

8 Chomsky 1965 on Deletion

Now consider:

(45)  a. At the reception, I met three boys who liked their professors and one girl.
b. At the reception, I met one boy who liked his professor and one girl.

Both of these sentences have sloppy interpretations, even though the deleted pronoun is not identical to the pronoun in the antecedent. For example, (45b) is represented as follows:

(46) …one boy who1 liked his1 professor and one girl <who1 liked her1 professor>

The problem that this sentence poses for condition (17) is that the two pronouns do not have the same phi-features values, and hence it looks like the relative clauses are not syntactically identical. The general issue was first noted by Ross (1986[1967]: 207), who defined the identity condition needed for deletion: “Constituents are identical if they have the same constituent structure and are identical morpheme-for-morpheme, or if they differ only as to pronouns, where the pronouns in each of the identical constituents are commanded by antecedents in the nonidentical portions of the phrase-marker.”
I propose to analyze facts such as (45) (and other cases of deletion where pronouns differ in phi-feature values) by applying a principle about deletion proposed by Chomsky 1965: 179 for the analysis of French examples like the following:

(47) ces hommes sont plus intelligent-s que Marie
these men are more intelligent-PL than Mary
“These men are more intelligent than Mary.”

Chomsky notes: “…the deletion of the Adjective of the embedded sentence should be blocked, since it differs from the Adjective of the matrix sentence in gender and number.” From such sentences, Chomsky draws the conclusion that “…the features added to a formative by agreement transformations are not part of the formative in the same sense as those which are inherent to it or as those which it assumes as it enters a Phrase-marker…. in the case of Adjectives and the copula (also Verbs, which take part in similar rules) the inflectional features that are added by agreement transformations are apparently not considered in determining whether the item in question is strictly identical with some other item.”

Whereas Chomsky’s principle only concerns inflectional phi-features on the copula and on adjectives, in (46) the relevant phi-features are those of pronouns. In order to apply Chomsky’s principle to (46), I assume following Collins and Postal 2012: 155, that the phi-features of pronouns are derived by agreement with an antecedent. Under that approach, at the point of relative clause deletion, the two pronouns in (46) do not differ in phi-features, since they do not have any phi-features at all.

In light of the analysis of pronouns given in Collins and Postal 2012, consider the following sentence:

(48) John met the boy who knows you, not the girl.

Collins and Postal assume that the antecedent of 2\textsuperscript{nd} person pronouns is a null DP in the left periphery that they call ADDRESSEE. Similarly, the antecedent of 1\textsuperscript{st} person pronouns is a null DP in the left periphery that they call AUTHOR. The phi-feature values of 1\textsuperscript{st} and 2\textsuperscript{nd} person pronouns are derived by agreement with these indexical DPs. If phi-features are derived by agreement, then what prevents the following representation?

(49) John met the boy who knows you\textsubscript{1}, not the girl <who knows me\textsubscript{2}>

The pronouns you and me in (49) are not co-indexed, and hence the deleted relative clause has no syntactically identical antecedent. Therefore, (49) violates (17) and (19).

9 Structure of Relative Clauses

Lastly, consider the consequences of relative clause deletion for the structure of relative clauses. On the traditional approach, a relative clause modifies a NP. The structure of the relative clause in (50) is given in (51):

(50) I saw the boy who I know.
On this analysis, relative clause deletion is simply deletion of the CP. However, formulation of relative clause deletion within the head raising analysis of Kayne 1994 runs into a problem. The structure of the relative clause in (50) according to Kayne 1994:90 is the following:

In this structure, the Spec of CP is a DP, headed by the D who. The NP moves from the complement of that DP to the specifier. The problem for (52), from the point of view of relative clause deletion, is that the string who I know is not even a constituent. Combining this with the assumption that only constituents can delete, structure (52) wrongly predicts that there should be no relative clause deletion.

I propose a slightly different version of the head raising analysis, which allows for relative clause deletion. In this analysis, the NP boy moves into Spec RelP.
In this structure, CP can be deleted to the exclusion of the raised noun. I will not be able to justify this structure any further here.

10 Conclusion

In this paper, I have investigated relative clause deletion and illustrated the consequences it has for the structure of relative clauses (section 9).

It remains to revisit other areas in the syntax of relative clauses and DPs from the perspective of relative clause deletion: stacked relative clauses, amount relative clauses, infinitival relative clauses, non-restrictive relative clauses, relative clauses with pied-piping, reduced relative clauses, other post-nominal modifiers, pre-nominal modifiers of various sorts, etc.

Comparative work on relative clause deletion will also be of great interest. For example, do prenominal relative clauses undergo deletion? The “relative deletion” discussed by van Craenenbroek and Lipták (2006) is different from the relative clause deletion discussed in this paper. Rather, van Craenenbroek and Liptak claim that the “relative deletion” is a kind of sluicing where the IP of a relative clause is deleted. In the data discussed in this paper, the whole relative clause is deleted. Whether or not relative clause deletion, in the sense of this paper, is available in any other language has yet to be investigated.

Acknowledgements:

I would like to thank Richard Kayne, Stephanie Harves, Jason Merchant, Paul Postal and Andrew Radford for useful discussion of the ideas and data in this paper. I also thank the students of Syntax II (Spring 2014, NYU) where I gave a lecture on the material.

References


Children manifest linguistic competence equivalent to that of adults in the first few years of life. By the age of three, children around the globe can produce and understand novel sentences, they can judge sentences to be true or false, and they can discern the inferences that sentences generate. Children’s experience dramatically underdetermines the linguistic abilities they so rapidly achieve, even given optimistic assumptions about children’s capacities to extract statistic regularities and form generalizations based on the input from adult speakers of the local language. The observation that children acquire a rich and complex grammar in the absence of abundant, if any, relevant experience forms the basis for poverty-of-the-stimulus arguments. These arguments and the experimental investigations that were inspired by them over the past fifty years have convinced many language scientists that young children know too much about human languages to have plausibly learned what they know from experience. As Chomsky (1965, p. 58-59) put it in Aspects:

It seems plain that language acquisition is based on the child’s discovery of what from a formal point of view is a deep and abstract theory – a generative grammar of his language – many of the concepts and principles of which are only remotely related to experience by long and intricate chains of unconscious quasi-inferential steps.

There is surely no reason today for taking seriously a position that attributes a complex human achievement entirely to months (or at most years) of experience … Such a position is particularly implausible with regard to language, an aspect of the child’s world that is a human creation and would naturally be expected to reflect an intrinsic human capacity in its internal organization.
Concrete demonstrations of children’s linguistic knowledge in the absence of experience have been reported in many experimental studies of children acquiring typologically distant languages. There is now substantial experimental evidence that innate linguistic principles govern children’s behavior as soon as they can be experimentally tested, by the age of three.

Universal Grammar (UG) is a theory of the initial state of language learners. The principles of UG determine the kinds of analyses that child language learners can adopt, as well as certain language specific knowledge that children bring to the task of language acquisition. As Chomsky (1965, pp. 25-27) remarks:

As a precondition for language learning, [the child] must possess, first, a linguistic theory that specifies the form of the grammar of a possible human language […].

[…] The important question is: What are the initial assumptions concerning the nature of language that the child brings to language learning, and how detailed and specific is the innate schema [… ] that gradually becomes more explicit and differentiated as the child learns the language?

Ten years later, Chomsky famously proposed that Universal Grammar restricts children’s grammars to structure-dependent hypotheses. This was empirically verified by Crain and Nakayama (1987), using the formation of English Yes/No Questions as a case study. As early as they could be tested, children produced structure dependent questions, and no instances of structure-independent questions were in evidence.

Structure-dependent linguistic principles dictate what sentences cannot be formed and what meanings cannot be assigned to sentences, so such principles are appropriately referred to as CONSTRAINTS on grammar formation. Constraints have come to play a central role in the nativist account of language development. Prior to the introduction of constraints into linguistic theory, grammars consisted of rules. Rules are positive statements, indicating which forms and meanings are possible in a language. In contrast to rules, constraints are negative statements, preventing child language learners from producing illicit forms and from assigning illicit meanings. In order for children to learn a constraint, they would require access to negative evidence or some substitute for it. Despite valiant attempts by researchers to identify sources of negative evidence, it is generally conceded that negative evidence is not available in sufficient quantity or at the right time to explain children’s rapid and effortless mastery of language (e.g., Bowerman 1988).

About a decade after the publication of Aspects, constraints began to be investigated by researchers in child language. Many of these researchers were convinced that “the rapidity and uniformity of language learning” could only be explained by universal, innately specified linguistic principles/constraints. Chomsky referred to these principles as “deep-seated regularities” that are “only remotely related to experience” (Chomsky 2006). We will illustrate one of these deep-seated regularities - Principle C of the Binding Theory. Principle C is one of the most studied linguistic constraints in child language. Principle C prevents children from assigning illicit meanings to several kinds of linguistic expressions.

Before we look at sentences that are governed by Principle C, it is instructive to consider the kinds of sentences that are not governed by Principle C, taking as examples the declarative sentence (1) and the wh-question (2).

(1) Obama thinks he will buy a Prius.
(2) Who thinks he will buy a Prius?

Both examples are ambiguous. The ambiguity turns on the interpretation of the pronoun he. On one interpretation, the pronoun he in (1) can be anaphorically linked to the name Obama, and the pronoun he in (2) can be anaphorically linked to the wh-word who. On this interpretation, (1) can be paraphrased as Obama thinks that he, Obama, will buy a Prius, and (2) can be paraphrased as Which person x is such that x thinks that x will buy a Prius? Both examples have a second interpretation according to which the pronoun he picks out some salient male individual who is not mentioned in the sentence. On this interpretation, the pronoun he is said to have direct (deictic) reference.

Now we can consider sentences that are governed by Principle C, such as examples (3) and (4). These sentences are not ambiguous, due to Principle C. In both examples, the pronoun he must receive direct (deictic) reference. In contrast to examples (1) and (2), the pronoun he cannot be anaphorically linked to the name Obama in (3) or to the wh-word who in (4).

(3) He thinks Obama will buy a Prius.
(4) Who does he think will buy a Prius?

In addition to declarative sentences and wh-questions, Principle C applies to discourse sequences. To see this, notice that the pronoun he used by Speaker A cannot be anaphorically linked to the name Obama in the response by Speaker B in (5). Generative linguists have suggested that Speaker B’s fragment answer involves copying and then deleting some of the linguistic material from Speaker A’s statement (e.g., Merchant 2004). This enables the Principle C constraint to operate on the underlying representation in Speaker B’s fragment answer, in which the pronoun he is present, but not pronounced.

(5) Speaker A: I know who he thinks will buy a Prius.
    Speaker B: Me too. Obama.

A number of experimental studies have been conducted with children as young as 3-years-old to see how young children interpret sentences like those in (1)-(5) (e.g., Crain and McKee 1986; Crain 2012; Crain and Thornton 1998; Thornton 1990; Conroy and Thornton 2005; Guasti and Chierchia 1999/2000; Kiguchi & Thornton 2004, Kiguchi & Thornton, in press). Essentially, children consistently adhere to Principle C. The findings from these experimental studies pose a challenge for the experience based account of language development, because the findings suggest that children do not avail themselves of either positive or negative evidence in the acquisition of constraints such as Principle C. On the other hand, the findings provide prima facie evidence for the nativist account of language development, as children evidently have linguistic knowledge for which there is no decisive evidence in the input.

2 Word Order, Scope and the Initial State

There is compelling evidence that children acquiring different languages begin at the same starting point, in keeping with Chomsky’s precept that “the innate schema … gradually becomes more explicit and differentiated as the child learns the language.” We will first illustrate one of the
different end-points that child language learners can reach. Then we will show that children, across languages, start off on the same path.

The phenomenon involves scope ambiguity, so a brief introduction to this topic is in order. The English sentence in (6) _All airplanes do not carry pets_ contains two logical expressions; one is the universal quantifier _all_ and the other is the negation marker _not_. As is often the case in sentences with two logical expressions, (6) gives rise to a scope ambiguity. On one reading, the quantifier _all_ takes scope over _not_. A paraphrase of this interpretation is ‘No airplanes carry pets.’ On the other (intended) interpretation, the negative marker _not_ takes scope over the universal quantifier _all_. A paraphrase of this interpretation is ‘Not all airplanes carry pets.’

(6)  _All airplanes do not carry pets._

Sometimes one interpretation of a scope ambiguity is strongly favored in a class of languages, whereas the alternative interpretation is strongly favored in another class of languages. Consider the English sentence (7).

(7)  _Ted didn't order pasta or sushi._

Sentence (7) contains two logical expressions, negation and disjunction (English _or_). One interpretation is clearly accessible to speakers of English. On this interpretation, negation takes scope over disjunction (NOT > OR). Under this scope assignment, there are two entailments, just as in one of de Morgan’s laws of propositional logic: \( \neg(A \lor B) \Rightarrow \neg A \land \neg B \). In example (7), _Ted didn't order pasta or sushi_, one of the entailments is that Ted didn’t order pasta, and the second entailment is that Ted didn’t order sushi. Taken together, this is called the ‘conjunctive’ entailment of disjunction in negative sentences. Disjunction generates a conjunctive entailment in (7) because negation takes scope over disjunction both in the surface syntax and at the level of semantic interpretation. Disjunction generates a conjunctive entailment in simple negative sentences in German, French, Greek, Romanian, Turkish, Bulgarian, and Korean.

In other languages, the surface syntactic position of disjunction and negation does not determine scope assignments. Example (8) is the Mandarin translation of the English sentence (7). Notice that word order (and surface syntax) is the same in Mandarin as in English, where the Mandarin word from negation is _méiyōu_, and the word for disjunction, _huòzhě_. Nevertheless, adult speakers of Mandarin do not judge (8) to license a conjunctive entailment. Rather, (8) means that either Ted didn’t order pasta _or_ Ted didn’t order sushi.

(8)  (Wǒ cāi) Tàidé _méiyōu_ diăn yídālimiànshì _huòzhě_ shòusìī.
    (I guess) Ted _not_ order _pasta_ _or_ _sushi._
    (我猜) 泰德没有点意大利面食或者寿司。
    ‘It’s either pasta or sushi that Ted did not order’

Although negation (_méiyōu_) takes scope over disjunction (_huòzhě_) in the surface syntax in (8), disjunction takes scope over negation at the level of semantic interpretation, at least for Mandarin-speaking adults. Other languages that favor the ‘inverse’ scope assignment include Japanese, Hungarian, Russian, Portuguese, Serbo-Croatian, Slovak, and Polish.\(^1\)

\(^1\) It might appear that these languages fail to conform to the relevant law of propositional logic: \( \neg(A \lor B) \Rightarrow \neg A \land \neg B \).
As a matter of fact, word order is completely orthogonal to scope assignment in human languages (Crain 2012). To see this, we need to identify two languages that differ in word order, as compared to Mandarin and English, but which also differ in scope assignment, just as English and Mandarin do. Japanese and Turkish fit the bill. In contrast to English and Mandarin, disjunction precedes negation in both Japanese and Turkish, as examples (9) and (10) illustrate.

(9)  Ted-wa pasuta ka sushi-o chu'umon shi-nakat-ta.
    Ted-TOP pasta or sushi-ACC order do-not-PAST
    ‘It’s either pasta or sushi that Ted did not order’

(10) Teddy makarna yada sushi siparis et-me-di.
    Teddy pasta or sushi order make-not-PAST
    ‘Teddy did not order either pasta or sushi’

Here is where things get interesting. Adult speakers of Japanese assign the ‘inverse’ scope interpretation to (9), as do adult speakers of Mandarin. Moreover, adult speakers of Turkish assign the ‘surface’ scope interpretation to (10), as do adult speakers of English. This indicates that word order is orthogonal to scope assignment in human languages.

These cross-linguistic differences in the assignment of scope relations to logical expressions have led to significant predictions about the course of language development. As Chomsky remarks in Aspects (p. 27):

[T]he main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning.

There is considerable evidence that children acquire human languages rapidly and uniformly. Nevertheless, languages vary in their properties and it takes some months (sometimes even years) for children to become aware of those properties that distinguish the local language from languages spoken in other linguistic communities. This invites the inference that, before children have mastered the intricacies of the local language, all children speak the same human language. If so, the language spoken by many children will differ from the language spoken by adults. These predictions were investigated in a series of studies of children’s assignment of scope relations across languages (see Crain 2012 for a review).

The first of this series of studies involved negated disjunctions. In an insightful study, Goro (2004) predicted that all children would initially assign the same scope relations in negative sentences with disjunction, regardless of the linguistic input they encountered. The reasoning is based on the observation that the different scope assignments for negated disjunctions across languages stand in a subset/superset relation. The scope assignment preferred by speakers of English is NOT > OR (i.e., negation takes scope over disjunction). This scope assignment

However, appearances are deceiving. Because disjunction takes scope over negation in these languages, negation does not influence the interpretation of disjunction. Disjunction is assigned the same interpretation in negative sentences as it is in affirmative sentences, and is subject to the same implicature of ‘exclusivity.’ For a review of a rich body of evidence that all human languages adhere to certain laws of first order logic, see Crain (2012).
asymmetrically entails the one that is preferred by speakers of Mandarin Chinese, OR > NOT. Based on this asymmetry, Goro reasoned that children would confront a potential learnability dilemma if they initially selected the superset scope assignment, OR > NOT. In the absence of negative evidence, children who initially assign this scope assignment would consistently have their hypothesis confirmed, and would fail to retract to the subset scope assignment manifested in languages like English, such that children acquiring these languages would fail to converge on the scope assignment in the local language. Since all children rapidly converge on the target language, Goro predicted that children acquiring all languages would initially assign the (subset) scope relations manifested in English.

Adopting the Principles and Parameters framework, Goro (2004) proposed that scope assignments were governed by lexical parameters. In the case of disjunction, Goro proposed that one value of the operative parameter forces disjunction to take scope over negation, as in Japanese and Mandarin, whereas negation takes scope over disjunction on the other parameter value. Adopting different terminology, Goro’s proposal was that disjunction words were Positive Polarity Items (PPI) in some languages but not in others. By definition, a Positive Polarity Item must take scope over (local) negation at the level of semantic interpretation, regardless of the structural relations that obtain between disjunction and negation in the surface syntax. Setting details aside, we can summarize Goro’s proposal as follows: disjunction words are associated with a lexical parameter, such that words for disjunction are either [+PPI] or [−PPI]. This led Goro to predict that children acquiring all human languages would initially assign a default (‘subset’) value to the lexical parameter, taking disjunction words to be [−PPI]. If children initially adopt the [−PPI] setting of the lexical parameter, then children acquiring ‘superset’ languages in which words for disjunction are [+PPI] will encounter adult input informing them to abandon the default ‘subset’ value [−PPI].

As an empirical consequence, adopting the [−PPI] value of the disjunction parameter would mean that Mandarin-speaking children should interpret the negated disjunction in (9), Tàidé méiyóu dìán yídàlimiànshí huòzhě shòusí, in the same way as English-speaking children and adults interpret the negated disjunction in (8) Ted didn’t order pasta or sushi. That is, Mandarin-speaking children should initially assign the scope relations designated as NOT > OR. This semantic interpretation is not attested in the input to Mandarin-speaking children, because adult Mandarin-speakers adopt the [+PPI] value of the disjunction parameter, hence adult speakers of Mandarin interpret negated disjunctions to have the opposite scope relations, OR > NOT.

The predictions generated by Goro (2004) have been pursued in seven languages so far. The chart below summarizes the findings. In four languages (Japanese, Mandarin Chinese, Turkish, and Russian), adult speakers assign the [+PPI] ‘superset’ value of the disjunction parameter. So, adult-speakers of these languages consistently accepted negated disjunctions corresponding to the English sentence Ted didn’t order pasta or sushi in circumstances in which Ted ordered only sushi, or only pasta. In contrast to adults, children acquiring these four languages rejected the sentences in the same circumstances, as did both child and adult speakers of languages that adopt the default value of the disjunction parameter [−PPI] (English, German, and Korean).

---

2 Consider ambiguous sentence S, with two possible interpretations, A and B. If interpretation A asymmetrically entails B, then A is true in a subset of the circumstances that make B true. A is the ‘subset’ interpretation, and B is the ‘superset’ interpretation.
Clearly children acquiring [+PPI] languages did not select the [–PPI] value of the disjunction parameter based on input from adults. According to Goro (2004), children’s initial selection of the –PPI value is due to a principle of language learnability called the Semantic Subset Principle (Crain, Ni and Conway 1994; cf. Berwick 1985). The Semantic Subset Principle (SSP) enforces an ordering on the values of parameters in cases where one value makes a sentence true in a subset of the circumstances that make it true on the other value. The SSP enjoins children to initially adopt the subset value of the parameter.

3 The Continuity Hypothesis

We have seen that child language can differ from the language of adults in the same linguistic community. Two accounts of such differences have been offered, corresponding to the two different perspectives on language acquisition - the experience-based approach, and the approach outlined by Chomsky in Aspects. We discuss these in turn.

According to the experience-based approach, the differences between child and adult language is a matter of degree. Child language is a constructive process. It develops gradually, based on readily available and salient cues in the input to children. Children use these cues to form generalizations, based on general purpose learning mechanisms including distributional analysis, analogy, and ‘cut and paste’ operations. The products of these learning mechanisms are 'shallow' piecemeal records (a.k.a. constructions/ templates/ schemas) of the linguistic patterns that are displayed in the children’s linguistic experience. Constructions are concatenated sequences of words, combined with category labels (NP, V, neg, INF, P) drawn from a simple inventory. Constructions are learned solely from positive evidence (Goldberg 2003, 2006; Pullum and Scholz 2002). When children’s generalizations extend beyond their experience, this is just an instance of a completely general problem of induction that arises for all learning that involves projecting beyond one’s experience (Cowie 1999). It is not necessary to suppose that children are innately endowed with "specific contingent facts about natural languages" (Pullum and Scholz 2002). If the data available to children are rich enough for them to determine the structures of human languages, given the right inferential techniques, then appeals to innately specified principles are unnecessary.
Adopting the alternative approach formulated by Chomsky in *Aspects*, there have been several specific proposals about the ways in which child and adult languages can differ. One proposal is called the Continuity Hypothesis (Crain and Pietroski 2001, Crain 2002, cf. Gleitman 1981, cf. Goodluck 1991, cf. Pinker 1984). According to the Continuity Hypothesis, child and adult languages can differ only in the ways that adult languages can differ from each other. The Continuity Hypothesis anticipates that children will try out various linguistic options, encoded in parameters, before they settle on the values of those parameters that distinguish the local language from other languages. In other words, children are expected to speak fragments of ‘foreign’ languages, from the perspective of adults. At any given time, children are speaking a possible human language, just not necessarily the language that is being spoken to them. The Continuity Hypothesis is, therefore, a rich source of poverty-of-the-stimulus arguments.

Experimental studies of child language have uncovered a number of instances in which children exhibit constructions that are not attested in the local language but are attested in other languages. We witnessed one example earlier. In resolving scope ambiguities, children acquiring Turkish, Russian, Mandarin and Japanese initially interpret negated disjunction in the same way as child and adults speakers of German, Korean, and English. This section reports the findings of two other studies in which child language has been found to differ from that of adults, but just in ways that adult language languages differ from each other.

The first example concerns the interpretation of negative statements with the Italian (deontic) modal expression *potere*. This example is interesting because Italian-speaking children appear to be speaking a fragment of English. Consider example (11).

(11) Gianni può non venire.

‘Gianni may not come’

Adult speakers of Italians judge (11) to mean ‘Gianni doesn't have to come’ (i.e., Gianni can come, but he is also allowed not to come’). Based on learnability considerations, Moscati and Crain (2014) predicted that Italian-speaking children would initially enforce the strong (‘impossibility’) meaning. More specifically, Moscati and Crain argued that sentences with deontic modals and negation are governed by a lexical parameter, R. The initial value for children should be the positive value [+R] for Italian-speaking children, whereas the parameter has a negative value [-R] for Italian-speaking adults. If so, Italian-speaking children and adults are expected to interpret sentences like (11) differently. Children should interpret sentence (11), *Gianni può non venire*, to mean that Gianni must not come (‘It is impossible/forbidden for Gianni to come’).

This was exactly what Moscati and Crain found. Whereas the scope relations assigned by adult speakers of Italian is determined by the relative positioning of modals and negation in the surface syntax, child Italian-speakers strongly prefer the inverse scope relations, as dictated by the [+R] value of the lexical parameter. The findings from the Moscati and Crain study, therefore, rule out any structure-independent account of children’s initial assignment of scope relations. The findings also speak against the possibility that children base their initial scope interpretations on evidence provided by adults. Italian-speaking children could not learn to assign the ‘impossibility’ meaning to (11), because adults cannot interpret (11) in this way. If children do not learn to assign the impossibility meaning, then the default [+R] setting of the lexical parameter is a viable candidate for innate specification. The final observation is that Italian-speaking children apply an interpretive principle that is operative in languages like
English, but not in Italian. More specifically, children’s non-adult interpretation of (11) is the same as the interpretation assigned by English-speaking children and adults to the sentence *Gianni can not come*. Therefore, Italian-speaking children are conforming to the natural seams of human languages, as anticipated by the Continuity Hypothesis.

The second example involves the acquisition of syntax. The conclusion is that some English-speaking children ask certain kinds of questions in the same way as Italian-speaking children and adults, but not in the same way as English-speaking adults. The questions under consideration begin with *wh*-words (*why, what, where, who*), so they are referred to as *wh*-questions. In adult English, *wh*-question words must be immediately followed by an inflected auxiliary verb (e.g., *be, do, can, have*). This yields well-formed questions like (12), with inversion of the subject NP (*you*) and the inflected verb (*are*). By contrast, *wh*-questions without inversion such as (13) are unacceptable, as indicated by the symbol ‘**’.

(12) Why are you here?
(13) *Why you are here?

Despite their unacceptability for adult English speakers, English-speaking children produce matrix (one clause) *why*-questions like (13) (e.g., Labov & Labov 1978). Based on an analysis by Rizzi (2001), Thornton (2004; 2008) proposed that English-speaking children’s matrix non-adult *why*-questions draw upon structural properties that distinguish English questions from those produced by speakers of Italian (also cf. de Villiers et al., 1990).

In Italian, the *wh*-word *perché* corresponds to English *why*. Italian *perché*-questions allow a grammatical option that is not permitted for the corresponding English *why*-questions. More specifically, a subject NP (e.g., *i tuoi amici*) is permitted to intervene between *perché* and the inflected verb (e.g., *hanno*) in Italian, so the question in (14) is acceptable, as is its counterpart with inversion (*perché hanno …*). The critical observation is that Italian permits matrix *perché*-questions that are similar in form to the non-adult matrix *why*-questions produced by English-speaking children, as we illustrated in (13) *Why you are here?*

(14) Perché i tuoi amici hanno finite il lavoro?
Why the-PL your friends have-3PL finished the-SG work
‘Why have your friends finished the work?’

Now consider the *perché*-question in (15). This is a ‘complex’ *wh*-question, with an embedded clause. Example (15) is ambiguous. On one interpretation, *perché* is construed locally. On this interpretation, the question asks about the reason for the event of "saying." A second ‘long-distance’ interpretation is also available. On this interpretation, the question asks about the reason for the event of “resigning.” This long-distance interpretation is only possible if the inflected verb (*ha*) is adjacent to the *wh*-word *perché*. This constraint on interpretation can be illuminated by comparing the question in (15) to the one in (16), where the subject NP (*Gianni*) intervenes between the *wh*-word *perché* and the inflected verb (*ha*). In contrast to (15), a long-distance reading is not possible in (16); the question only asks about the reason for the event of “saying,” not about the event of “resigning.” As indicated by the gloss below (16), the same local construal is unequivocal in English *how-come* questions such as *How come Gianni said he would resign?* In short, the forms taken by matrix *why*-questions differ across languages, but the long-distance construal of complex *why*-questions is governed by a linguistic constraint.
(15) Perché ha detto che si dimetterà?
  why have-3sg said that self resign-3sg/future
  'Why did he say that he would resign?'
(16) Perché Gianni ha detto che si dimetterà?
  why Gianni have-3sg said that self resign-3sg/future
  'How come Gianni said that he will resign?'

This difference between matrix and complex why-questions was exploited in a study by Thornton (2008). The goal of the study was to explain English-speaking children’s non-adult matrix why-questions. Adopting the Continuity Hypothesis, Thornton (2008) proposed that English-speaking children initially analyze why-questions in the same way as Italians analyze perché-questions. To assess this possibility, Thornton conducted an extensive experimental and longitudinal diary study of the wh-questions produced by one child, AL, between the ages of 2;0 and 5;6 years. By the age of 3;4, all of AL’s matrix wh-questions were adult-like, except for why-questions. Over a year after that, the majority (80%) of AL’s matrix why-questions continued to be non-adult. Examples of AL’s adult-like and non-adult wh-questions are given in (17) and (18).

(17) a. How did Tweetie get maked? (3;0)
    b. When will we big enough to climb up there? (3;1)
    c. Who was it that you was talking to? (3;1)
(18) a. Why the pig got eatened? (3;0)
    b. Why unicorns are pretend? (3;1)
    c. Why you are going in that one? (3;2)

The critical data for evaluating the Continuity Hypothesis were AL’s complex wh-questions with a long-distance interpretation. As we have seen, the long-distance interpretation of complex wh-questions, including why-questions, is governed by a linguistic constraint that requires the inflected verb to be adjacent to the wh-phrase. AL clearly adhered to this constraint. Between the ages of 3;0 and 5;6, AL produced 83 complex wh-questions (with think and say as the matrix verb). Twenty-one of these wh-questions contained why, and 62 contained other wh-words. All of AL’s wh-questions with wh-words other than why were adult-like. Moreover, seventeen of AL’s 21 complex why-questions were adult-like, with inversion of the inflected verb (in the form of do-support). AL’s adult-like complex why-questions were long-distance wh-questions, as is evident in the examples in (19).

(19) a. Why do you think Santa’s not coming this year? (3;10)
    b. Why do you think Boomer came in with us? (4;2)

To recap, an intensive investigation and analysis of the wh-questions produced by AL support the conclusion that English-speaking children’s non-adult why-questions are derived in the same way as perché-questions in Italian. In producing non-adult why-questions, English-speaking children ignore abundant evidence that indicates a mismatch between their why-questions and those of adults. If AL’s productions are indicative, the long-distance why-questions produced by English-speaking children should always be adult-like. The disparity between AL’s non-adult matrix why-questions and adult-like complex why-questions is compelling evidence for
the Continuity Hypothesis and the principles of Universal Grammar. This pattern is difficult to reconcile on an experience-based approach to language acquisition.

4 Conclusion

As Chomsky made clear in Aspects, a central task of linguistics is to explain how children bridge the gap — Chomsky calls it a ‘chasm’ — between what they come to know about language and what they learn from experience. Proponents of the experience-based approach to language development accuse nativists of overestimating the complexity of what children learn, underestimating the data children have to work with, and manifesting undue pessimism about children’s abilities to extract information and form generalizations based on the input. Nativists working within the generative tradition have championed an alternative argument - that children come to know linguistic facts that are “only remotely related to experience by long and intricate chains of unconscious quasi-inferential steps.” This brief report surveyed some representative examples of empirical finding from studies of child language that support the nativist approach advanced by Chomsky 50 years ago. In each case, children have been found to speak a fragment of a language that is spoken by adults who reside somewhere else in the world, but not by adult speakers of the local language. These examples constitute poverty-of-stimulus arguments, and lend credence to the Continuity Hypothesis which supposes that, at every stage of language acquisition, children speak a possible human language.

References


RAISING THE SUBJECT OF THE ‘OBJECT-OF’ RELATION

MARCEL DEN DIKKEN
CUNY Graduate Center

In section 2.2 of chapter 2 of Aspects, Chomsky (1965) stresses the inherently relational nature of grammatical functions, and proceeds by ‘defining “Subject-of,” for English, as the relation holding between the NP of a sentence of the form NP Aux VP and the whole sentence, “Object-of” as the relation between the NP of a VP of the form V NP and the whole VP, etc.’ (p. 69). The idea that the ‘object-of’ relation involves complementation remained standard until the late 1980s, when the possibility of modeling this relation in terms of specification became a serious possibility with the introduction of Larsonian ‘VP shell’ structures (Larson 1988) for ditransitive constructions as well as a range of other transitive sentence types — including ones in which the direct object is followed by a manner adverb (such as John read the letter carefully, with carefully mapped onto the complement position of the verb, and the letter projected as the specifier of the inner VP). Neither Larson (1988) nor Hale & Keyser (1993), who take the ‘VP shell’ programme further, propose to systematically represent the ‘object-of’ relation in terms of specification; but Bowers (2010:1) does: for him, ‘every argument is merged as the specifier of a functional category of a particular type’ (see also Lin 2001, Borer 2005, Lohndal 2012 and the many references cited there). With the functional categories introducing arguments taken to be the RELATORS of Den Dikken (2006), the ‘subject-of’ and ‘object-of’ relations can be unified under the general rubric of predication relations: the ‘object’ is the subject of the minimal VP predicate; the ‘subject’ is the subject of the larger constituent formed by the predication relation between the VP and the object; each predication relation is mediated by a RELATOR, the higher one of which is now customarily identified as v (Chomsky 1995:Chapter 4) or Voice (Kratzer 1996), as in (1). (Henceforth, ‘v/Voice’ = RELATOR1.)

(1) \([\text{RP}_1 \text{VP SUBJECT } \text{RELATOR}_1 = v [\text{RP}_2 \text{ OBJECT } \text{RELATOR}_2 [\text{VP V (...)])}]]]

With the ‘object-of’ relation exploiting the specifier position of a predicational RELATOR phrase immediately outside the minimal VP, the complement-of-V position becomes available for other material — secondary predicates, in particular. Both Larson (1988) and Hale & Keyser (1993) call upon the complement-of-V position for this purpose; but convincing arguments recommending their approach over the alternative ‘small clause’ analysis (according to which secondary predication relations are established wholly within a small clause, either in the complement of the verb or (in the case of depictives) in an adjunction position to the projection of the verb; see i.a. Stowell 1983, Hoekstra 1984, 1988, 2004 and references there) have not been forthcoming. I will begin my discussion here with an empirical argument
suggesting that the two approaches are not in competition with one another but should both be available for descriptive adequacy to be achieved.

A structure of the type in (2), where the complement-of-V position is occupied by the AP secondary predicate, seems to come in handy for resultative secondary predication constructions such as (3), which are often said to involve a thematic relation between the verb and the subject of the resultative secondary predicate: that is, it is often claimed that the floor is the internal argument of the verb, with the latter sometimes represented as a complex predicate (paint white; this goes back to Chomsky’s 1965:190 approach to verb-particle constructions). Such claims lend themselves naturally to a representation of (3) as in (2), where the VP is the complex predicate related to the object in SpecRP2.

(2) \[RP1=VP John[REL1=v [RP2 the floor [RELATOR2 [VP V [AP white]]]]]]
(3) John painted the floor white

Hoekstra (1988) points out, however, that an interpretation of sentences such as (3) in which the floor is the thematic object of the painting event is not the only one supported by such sentences. Though for English (3) it is generally reported to be difficult to access the other reading (for reasons that are not understood), its Dutch counterpart in (4) readily supports an interpretation according to which Jan is clumsily engaged in a painting event directed, for instance, at the ceiling that results in white paint being splattered all over the floor. And even for English speakers, resultatives can easily be found in which the ‘object’ is not the thematic object of the verb: (5) is ambiguous between a reading in which the painting event was directed at John’s hands and a more salient reading in which John arduously painted some unspecified object to such an extent that his hands got sore. Unsurprisingly (in light of the ambiguity of (4)), the Dutch equivalent of (5), given in (6), is ambiguous in the same way.

(4) Jan heeft de vloer wit geverfd
   Jan has the floor white painted
(5) John painted his hands black and blue
(6) Jan heeft zijn handen bont en blauw geverfd
   Jan has his fingers motley and blue painted

For interpretations of (3)–(6) on which the object is not the ‘paintee’, we do not want to assume a thematic relation between the verb paint and the object: rather, the latter is uniquely the subject of the secondary predicate, not an argument of the verb paint or the complex predicate paint+AP.

Intuitively, we would like the syntactic structure of resultatives with such interpretations to feature an implicit object in the thematic object position (i.e., the specifier position of RP2), representing the thing that is the ‘paintee’, and an overt subject just for the secondary predicate. The approach to the ‘object-of’ relation reflected by (1) affords us just such a structure:

(7) a. \[RP1=VP John[REL1=v [RP2 OBJ[REL2 [VP V [RP3 the floor [REL3 [AP white]]]]]]]]
   b. \[RP1=VP John[REL1=v [RP2 OBJ[REL2 [VP V [RP3 his hands [REL3 [AP black and blue]]]]]]]]

In (7), ‘OBJ’ is an implicit object — one that receives a thematic role from the complex predicate dominated by VP but does not and cannot receive a phonological matrix: the overt subject of the small clause in the complement of V sups up the only accusative case feature available in the structure.\(^1\)

\(^1\) This caseless unspecified object is the very same thing that also figures in the structure of sentences such as John was eating, for which Chomsky (1965:87) advocates a deletion approach (‘unspecified object deletion’, UOD). Rather than assuming deletion of an object, I adopt here an approach (couched within the ‘late insertion’ framework of Distributed Morphology) in which ‘OBJ’ receives no phonological realisation (because it lacks case it is not licensed to be spelled out). The caselessness of ‘OBJ’ is apparent from the syntax of Romance causative constructions with UOD. As Postal (1977) points out for French, while the ordinary transitive causative in (ia) forces the causee to be introduced by the dative preposition à, its UOD counterpart in (ib) has an accusative causee, facilitated by the fact that accusative case is not used up by the object of the causatived verb in this
Though resultative constructions have been discussed at considerable length in the generative literature, what has not, to my knowledge, been noted before is that the ‘unspecified object’ readings for sentences such as those in (3)–(6) are unavailable in their minimal counterparts in (8)–(11). The sentences in (8)–(11) are all grammatical but they only support readings in which the object is the thematic argument of the painting event. The particles\(^2\) that were added to (3)–(6) to yield (8)–(11) apparently force the physical object to occupy the specifier position of RP2, where it is interpreted as the subject of the complex predicate dominated by VP: while (2) is a grammatical structure for (8)–(11), (7) is not.

(8) John painted the floor *up* white  
    John repainted the floor white  
(9) Jan heeft de vloer wit \{*over geverfd/af gelakt*\}  
    Jan has the floor white over painted/off lacquered  
(10) John painted his hands *up* black and blue  
    John repainted his hands black and blue  
(11) Jan heeft zijn handen bont en blauw \{*over geverfd/af gelakt*\}  
    Jan has his hands motley and blue over painted/off lacquered

We can understand this if we assume that the additional particles in (8)–(11) are the occupants of the head of RP2: in (13), REL2, when realised as a particle, makes it impossible for a case-checking relation to be established between \(v\) (the source of accusative case) and *the floor* in SpecRP3.\(^3\)

(12) \([_{RP1=v}John [_{REL1=v} [_{RP2=OBJ} [_{REL2=PRP} [_{VP V [_{AP white}]]]]]]]\\)

With (13) excluded, the particle constructions in (8)–(11) must be derived along the lines of (12), where the object occupies the specifier position of RP2, and is hence interpreted as the subject of the complex predicate dominated by the VP.

For the complex particle constructions in (14)–(16) (see Kayne 1985 for seminal discussion), an analysis that treats the particle as the spell-out of REL2, outside the VP, is now a logical next step.

(14) a. they made John *out* a liar  
    b. they made John *out* to be a liar  
(15) a. John turned *out* a liar  
    b. John turned *out* to be a liar  
(16) a. John ended *up* a liar  
    b. John ended *up* being a liar

---

\(^{2}\) See Keyser & Roeper (1992) on the fact that the prefix re- in English distributes very much like particles.

\(^{3}\) The case-based approach follows in the footsteps of Den Dikken’s (1995) analysis of complex particle constructions. Technically, the ban on a case-checking relation across an overtly filled REL2 can perhaps be derived from a phase-theoretic approach, on the assumption that the “phase extension” (in the sense of Den Dikken 2007) necessary to render \(v\) and *the floor* local to one another fails when REL2 is filled by a particle. I will not spell out the details here; it may not be entirely straightforward to work them out. An alternative rationale for the ungrammaticality of (13) could be built on the assumption that an overt element in REL2 demands an overt element in its specifier. This alternative is less obviously syntactic in nature, and it also does not carry over very easily to some of the facts discussed later in this paper, so I will not adopt it here.
The small clause (John a liar) or TP (John {to be/being} a liar) is intuitively the internal argument of the verb+particle complex: what was made out, in (14), is that John is a liar; what turns out, in (15), is that John is a liar; and what we end up with in (16) is that John is a liar. But generating the small/non-finite clause in the structural object position, to the left of the particle, is problematic: its predicate (a liar) strongly resists being pronounced in that position (*they made John (to be) a liar out etc.).

The sentences in (14)–(16) are semantically equivalent to those in (17)–(19), with the proleptic (‘expletive’) pronoun it (originating in the SpecRP2 position) serving as a ‘place-holder’ for the propositional argument.4

(17) they made (it) out that John was a liar  
\[ \text{[RP1-VP they [REL1=v [RP2 OBJ=(it) [REL2=PRT [VP V [CP that John was a liar]]]]]} \]

(18) it turned out that John was a liar  
\[ \text{[TP it turned [RP2 OBJ=it [REL2=PRT [VP V [CP that John was a liar]]]]]} \]

(19) it ended up that John was a liar  
\[ \text{[TP it ended [RP2 OBJ=it [REL2=PRT [VP V [CP that John was a liar]]]]]} \]

In (17)–(19), it is linked to the that-clause in post-particle position — the two form an ‘expletive–associate CHAIN’ in the sense of Chomsky (1986). The pronoun in the examples in (18) and (19) must be overt (for EPP reasons), just as in the more familiar raising constructions with seem and appear; since the (classic, Chomsky 1981-style) EPP is not in effect for the SpecRP2 position, an overt pronoun is not required when there is no raising to the structural subject position, as in (17). But when it is not pronounced, a silent token of it still takes care of the association of the that-clause with the ‘object-of’ function, by proxy. I will turn shortly to the question of what happens to ‘OBJ’ in the examples in (14)–(16), and to the question of how the surface word orders of these examples are derived. But first, I would like to present a few simple empirical arguments in favour of VP-external base-generation of the particles in these examples.

One immediate advantage of treating the particle as the realisation of the VP-external REL2 is that the strings to be a liar and being a liar end up, by themselves, in the complement position of V. This is desirable in light of the idiosyncratic selectional relation between, on the one hand, turn out and a to-infinitive and, on the other, end up and an -ing gerund.5 The selection of a to-infinitive or a gerund is based on the verb, not on the particle: just like something ends being, rather than to be, boring, so it ends up being, rather than to be, such; I brought him up to be my pride and joy is fine, so there clearly is no problem with the particle up combining with a to-infinitive. Since selectional relations are highly local relations, what we need for (15) and (16) is a structure in which the non-finite clause is the complement of V, not separated from it by a particle. Structures of the type in (17)–(19) guarantee precisely this.

A second advantage of placing the particle outside VP is that it frees up space inside the VP for other functional elements for which it is plausible to treat them as spell-outs of RELATORS — such as as in (20) or voor ‘for’ in Dutch (21) (see Den Dikken 2006 on as and voor as RELATORS).

(20)  
a. they made John out as a liar  
b. John turned out as a serious candidate  
c. John ended up as a nice guy

(21) ze maakten Jan uit voor leugenaar  
they made Jan out for liar

---

4 Den Dikken (1995:54) states that it is impossible in (17), but as Larsen (2014) points out, this is inaccurate: there are many attested instances of made it out that S, and many native speakers (though apparently not all) find those acceptable.

5 While end up to be x seems much worse than end up being x, turn out being x is not infrequent (though much less common than turn out to be x)— see the discussion in the on-line forum http://forum.wordreference.com/showthread.php?t=2038538.
The underlined elements in (20) and (21) can now be treated as spell-outs of the head of an RP in the complement of V — a small clause in whose specifier position the object originates:

\[
(22) \quad [\text{REL}2=\text{PRT} \left[ \text{VP} \right. \left. \text{P} \left[ \text{RP} \text{ John} \left[ \text{REL}3=\left(\text{at} \left[ \text{sNP} \text{ a liar} \right] \right) \right] \right] \right]]
\]

Thirdly, base-generation of the particle outside the VP rather than within the complement of V yields an immediate pay-off in the account of the facts in (23)–(25), involving inversion of the predicate of RP3 around its subject and obligatory realisation of a copular element in the complement of V.

(23) they made the biggest liar out *(to be) John
(24) the biggest liar turned out *(to be) John
(25) the biggest liar ended up *(being) John

For familiar copular inversion cases such as (26b), the obligatoriness of the copula in a context in which (as (26a) shows) it is in principle optional is commonly accounted for with an appeal to a ‘space problem’. copular inversion is a movement operation manoeuvring the predicate around its subject, which requires for there to be space in the structure for the predicate to land in; the copula provides this space, and is therefore obligatory.

(26) a. I consider John *(to be) the biggest liar
    b. I consider the biggest liar *(to be) John

If the particles out and up in (23)–(25) originated within the complement of V, between V and the small clauses establishing the predication relation between the biggest liar and John (RP3 in (22)), the fact that these particles do not ‘solve the space problem’ would be difficult to account for. In (27)–(28) (due to Heycock 1994) and the double-object construction in (29) (analysed by Den Dikken 1995 in terms of predicate inversion), inversion is successful without the need for a copula precisely because in all of these constructions there is indeed a ‘space-maker’ in between the small clause and V: an aspectual head in (27)–(28) (see Den Dikken 2006:146–7 for discussion) and the particle out in (29) (see Den Dikken 1995). But apparently in (23)–(25) there is not: for predicate inversion to come out grammatical here, a token of the copula is required in (23)–(25) just as it is in (26b). This follows directly from the hypothesis that the particles in these examples originate outside VP, as depicted in the structures below the examples in (17)–(19) and in (22): in that high position, they cannot help make predicate inversion (which is a strictly local operation) possible, so copulas are called upon to bring this about.

(27) the best solution becomes/remains instant retreat
(28) if Bill has an alibi for 6 p.m., that makes the murderer John

---

6 I am presenting the problem in the simplest of terms here in order not to unduly complicate the discussion in this short paper; for technical details and references to relevant literature, see Den Dikken (2006).

7 I am following my 1995 analysis of ditransitive verb-particle constructions here, generating the particle in the complement of V. But for the constructions in (14)–(16) I now take a different approach from the one advocated in Den Dikken (1995). The heterogeneity of the new analysis is justified on semantic as well as syntactic grounds: while in *they sent a schedule out to the stockholders* there is an intuition that *out* forms a tight-knit semantic unit with *to the stockholders*, for *they made John out a liar* and its ilk there is no sense in which *out* belongs closely to the predicate of the small clause; and while extraction of the predicate *a liar* and modification of the particle are impossible and placement of the particle is fixed in (14)–(16), we find a very different tangle of facts for *they sent a schedule out to the stockholders* (see Den Dikken 1995 for discussion). The fact that *they sent the stockholders out a schedule* behaves in many ways very much like *they made John out a liar* is one that I cannot address explicitly here; but see Den Dikken (1995) for what I still consider to be a plausible analysis of the properties of the former.
(29) they sent the stockholders out a schedule
   \[ \text{VP } \text{V} = \{ \text{P} \text{ the stockholders}, \text{PRT} = \text{out} \} \text{ [RELATOR t]} \]}

With these three arguments cementing the foundations for the analyses of (14)–(16), we now face three important questions: (a) what about the implicit object ‘OBJ’ in the structures for these sentences? (b) how is John case-licensed (bearing in mind that we had blamed the ill-formedness of (13) on a failure on the part of the floor to get its case feature checked)? and (c) how do the structures of (14a,b) accommodate the output word order, with John to the left of the particle?

While in (17) the proleptic object pronoun is optionally overt, in (14) it cannot be spelled out as it at all: *they made it out John (to be) a liar is entirely impossible. This is directly parallel to the fact that in constructions with a small or non-finite clause in the complement of turn out or end up (recall (15)–(16)), the structural subject position cannot be filled by it: *it turned out John (to be) a liar,8 *it ended up John (being) a liar. The problem in all these cases is that the case-dependent noun phrase John cannot get its case feature checked: it usurps the only case feature that John might hope to check (accusative in the make out case, nominative with turn out/end up). But even when it is not present, John cannot freely be spelled out inside the complement of V: while *the made out John to be a liar is generally acceptable, *they made out John a liar is rejected by the vast majority of speakers.9 We had already established in the discussion of (13) that the particle blocks in-situ case checking for the subject of the small clause in the complement of V. For (13) there was no way out: raising of the floor into the ‘OBJ’ position would turn it into the subject of the VP predicate paint white, which would give the sentence an interpretation different from the one intended — the intended reading for (13) is one in which the floor is not the ‘paintee’; but raising the floor into SpecRP2 would deliver a reading in which it is. By contrast, in the case of (14), raising John into SpecRP2 is not semantically undesirable: John will be interpreted as the subject of the complex predicate make out a liar.10

The fact that *they made out John to be a liar is much better than *they made out John a liar is ascribed by Lasnik (1999, 2003) and Hong & Lasnik (2010) to a mysterious difference between to-infinitival and small-clause ECM subjects: while the latter are compelled to raise, the former do so only optionally. Not only is optionality of operations generally difficult to countenance in the minimalist era, there is no obvious theoretical rationale for a difference between to-infinitives and small clauses with respect to the obligatoriness of the raising of their subjects. It seems to me more likely that the grammaticality of make out x to be y is to be related to the fact that the infinitival clause of make out constructions can optionally be introduced by the complementiser for, and that, whenever for is present, the subject of the infinitival clause must perforce be to the right of the particle (they made out for John to be a liar).11

8 On a par with *it seemed/appeared/was likely John (to be) a liar, the canonical raising environment.
9 The judgements reported here on make out a liar constructions are those originally reported by Kayne (1985). See Larsen (2014) for a somewhat more nuanced perspective which does not, however, invalidate Kayne’s original characterisation of the empirical lie of the land.
10 In the case of (14a) evidence for raising of John into SpecRP2 as opposed to base-generating John in the SpecRP2 position is hard to come by. Floating quantifier distribution, for instance, does not confirm raising of John from a position to the right of the particle: he made the children <all> out <all> crazy is clearly much worse with all between out and crazy. On the other hand, in the corresponding to-infinitival construction, he made the children <all> out <all> to be crazy, placement of all to the right of out is natural — which suggests that in the to-infinitival case there definitely is raising going on, not surprisingly in light of the EPP (which requires SpecTP to be occupied). For the sake of uniformity, I will assume raising in both the small-clause and the to-infinitival case, which is perhaps supported on the basis of the fact that extraction from the subject of the small clause is entirely impossible in the make x out a liar [lang2057] construction: the strong deviance of *who did they make a friend of out (to be) a liar could be attributed to a freezing effect if the small-clause subject is forced to raise into SpecRP2 instead of being base-generated there.
11 Below are some attested examples of the for-to type, found on the internet:

   (i) a. they made out for you to be old
   b. worse than Gabe made out for him to be
   c. if the lake was really as bad as they made out for it to be
   d. I didn’t think it was as bad as they made out for it to be
This datum can be taken to lead us further to the hypothesis that whenever the subject of the infinitival clause follows the particle out, it is case-licensed inside the infinitival clause by the infinitival complementiser, which has a silent allomorph (see Kayne 1984 on ECM in want-constructions, and also further below). Optionality is now reduced to lexical choices (selection of a ‘bare’ infinitival TP or a CP; overt or silent allomorph of for), a natural point of variation.

The proposal, then, is that in they made John out to be a liar the infinitival’s subject raises into SpecRP2, a position in the matrix clause, whereas in (3)they made out John to be a liar the infinitival clause is a CP with a case-licensing C-head that keeps the subject inside its clause. How does this make sense of the scope facts that Lasnik (1999, 2003) and Hong & Lasnik (2010:281) present? They point out that make out x to be y permits narrow scope for every even number in (30a) (‘the mathematician determined that it is not the case that every even number is the sum of two primes’) and no students in (31a) (‘the FBI determined that there were no students guilty’), while the make x out to be y order in (30/31b) forces wide scope for every even number and no students ‘(the mathematician determined for every even number that it is not the sum of two primes’, ‘for no students is it the case that the FBI made them out to be guilty’). To complete the picture, I add that for the small-clause case in (31c), which Hong & Lasnik do not present, we get the same behaviour as for (31b): wide scope of no students is again the only option.

(30) a. the mathematician made out every even number not to be the sum of two primes
   b. the mathematician made every even number out not to be the sum of two primes

(31) a. the FBI made out no students to be guilty
   b. the FBI made no students out to be guilty
   c. the FBI made no students out guilty

The fact that (30b) and (31b,c) are scopally unambiguous and force wide scope for the quantifier can be understood from the point of view of the analysis proposed for these sentences once we realise that their derivations involve movement into a 0-position: the SpecRP2 position is the ‘object-of’ position, its occupant a subject of predication. Movement into a 0-position (heavily exploited in the ‘movement theory of control’) does not reconstruct: the thematic landing-site of movement is a position in which the moved element must be interpreted (cf. Hornstein 1998). This, then, is genuine ‘raising to object’ or ‘subject-to-object raising’ (Postal 1974), made possible by the representation of the ‘object-of’ relation as a specification relation.12

I suggested three paragraphs back that the linear order make out x to be y is derived from a structure in which make out takes an infinitival CP as its complement. How does this structure allow the quantifier in (31a) to take matrix scope? Here we should tread with care. Larson et al. (1997:89, fn. 37) point out that quantified subjects of want-type verbs, for which they assume, following Kayne (1984), that they take a CP complement, ‘do not routinely show the same scopal possibilities as quantified subjects of believe-type verbs, even when the complementizer for is absent’. Wide scope for every defendant vis-à-vis the matrix subject is harder to get in (32) than it is in (33); and within (32), it is harder in the a–sentence (which has overt for) than in the b–example. But even in (32a) it does not seem entirely impossible to get an inverse scope reading.

(32) a. some juror wants for every defendant to be acquitted
   b. some juror wants every defendant to be acquitted

(33) some juror believes every defendant to be guilty

The infinitival CP itself does not seem to be a strong impediment to the establishment of a dependency across it (as witness the grammaticality of he prefers for himself to win; see Larson et al. 1997:65, fn. 13). The difficulty of getting wide scope for every defendant in (32a) may ultimately have more to do

12 With objects analysed as complements, ‘raising to object’ is of course a complete anomaly in a minimalist theory of internal Merge: movement into complement position cannot extend the tree.
with the presence of an overt complementiser\textsuperscript{13} than with the opacity of the infinitival CP. Why it is still relatively difficult to obtain wide scope for every defendant when the C-head is silent (as in (32b)) is not entirely clear. I will assume it to be structurally possible, in (32b) as well as in (30a) and (31a).

To make the empirical picture complete, we should include the versions of (32b) and (33) lacking to be, given in (34) and (35). These both resist wide scope of the universal QP, behaving just like (31c) in allowing linear scope only (see Hong & Lasnik 2010 and earlier references cited there).

(34) some juror wants every defendant acquitted
(35) some juror believes every defendant guilty

For (31c) we have already derived its frozen scope, from movement into a θ-position (literal ‘raising to object’). For (34) and (35) we are now led to make the same assumption. But while for the make out construction the need to raise the small-clause subject was made to fall out from the hypothesis that the particle out obstructs a case-checking relation between v and the in-situ subject of guilty, there is no particle around in the examples in (34) and (35) to do the same thing. So why is ‘raising to object’ obligatory here as well?

Let me start with the contrast between (32b) (which allows inverse scope, albeit not quite as easily as does (33)) and (34) (where wide scope for the universal QP is unavailable). Following Kayne (1984) and Larson et al. (1997), I have assumed that in (32b) the infinitival clause is a CP with a silent complementiser — so the structure of (32b) is similar to that of (17), featuring a silent proleptic object ‘OBJ’ that serves as a place-holder for the CP in the complement position of V. In (36), the universal QP does not raise into the ‘object-of’ position, and is structurally allowed to take either wide or narrow scope with respect to some juror (assuming, as before, that silent infinitival C does not obstruct QR). For the corresponding small-clause example in (34), a structure identical to (36) save for the category of V’s complement, given in (37), is ill-formed. Small clauses cannot be associated with proleptic objects because of a featural non-match: the proleptic object is a pronominal element, equipped with φ-features; but small clauses, unlike CPs, have no φ-features hence cannot serve as associates of pronouns.\textsuperscript{14} With (37) ill-formed, we must resort to (38) instead. Here ‘raising to object’ takes place, freezing the universal QP’s scope. This derives the lack of ambiguity in (34).

\begin{equation}
(36) \quad [\text{RP1-VP} \text{ some juror } [\text{REL1-v [RP2 OBJ]} [\text{REL2 [VP V [CP C every defendant to be acquitted]]]}]]
\end{equation}

\begin{equation}
(37) \quad [* [\text{RP1-VP} \text{ some juror } [\text{REL1-v [RP2 OBJ]} [\text{REL2 [VP V [RP3 every defendant acquitted]]}}]]]
\end{equation}

\begin{equation}
(38) \quad [\text{RP1-VP} \text{ some juror } [\text{REL1-v [RP2 every defendant] [REL2 [VP V [RP3 it acquitted]]}}]]
\end{equation}

With the account of (34) now in place, the scopal rigidity of (35) is no longer a mystery: the derivation of (35) is entirely on a par with that of (34), given in (38). The question now is a different one: how can we avoid ‘raising to object’ in the example in (33)? If the structure of (33) involves an infinitival TP in the complement of V, as in (39), it will be extremely difficult to prevent ‘raising to object’: the infinitival TP is just like the small clause in (37) in lacking φ-features,\textsuperscript{15} hence it should not be able to serve as the associate of a proleptic pronoun in the object position; the structure in (39) is ungrammatical. To get a grammatical syntax for (32) we can do either of two things: (a) we can perform ‘raising to object’ on the subject of the infinitival TP in the complement of V, as in (40), or (b) we can postulate a null-headed CP in the complement of V, associated with a proleptic pronoun in the ‘object-of’ position, as in (41).

\textsuperscript{13} Apparently not subject to deletion at LF, pace Lasnik & Saito’s (1984) discussion of that.

\textsuperscript{14} Their lack of φ-features also accounts for the fact that they cannot serve as subjects of finite clauses. See Safir (1983) and Den Dikken (1987) on cases like workers angry about the pay was just the sort of situation that the ad campaign was designed to avoid, involving what Safir calls ‘Honorary NPs’, a term that should be taken quite literally and be translated into an analysis in which the small clause [workers angry about the pay] is dominated by a noun phrase with a silent head (something like ‘SITUATION’). Small clauses by themselves cannot be subjects.

\textsuperscript{15} Just like small clauses (recall the previous footnote), infinitival TPs cannot serve as subjects.
Raising the Subject of the ‘Object-of’ Relation

(39) \(*_{LP} \text{some juror} [\text{REL}=v_{LP} \text{OBJ}, \text{REL}_2 [\text{VP} \text{V} \text{TP every defendant to be guilty}]]\)\]
(40) \([_{LP} \text{some juror} [\text{REL}=v_{LP} \text{OBJ}, \text{REL}_2 [\text{VP} \text{V} \text{TP every defendant to be guilty}]])\]
(41) \([_{LP} \text{some juror} [\text{REL}=v_{LP} \text{OBJ}, \text{REL}_2 [\text{VP} \text{V} [\text{CP C every defendant to be guilty}]]])\]

The PF outputs of the structures in (40) and (41) are indistinguishable. But their LFs are not — while (40) blocks wide scope for the universal QP, (41) allows this. Of course the narrow-scope reading for every defendant that both (40) and (41) support is readily available for (33). But (33) also supports an inverse scope interpretation, and that interpretation cannot be obtained via the ‘raising to object’ derivation if what I argued above is accurate. There is another thing that this derivation does not support: sentences with expletive there or idiom-chunk ECM subjects, which both resist the 0-role assigned in the SpecRP2 position, should not be derivable via ‘raising to object’. The fact that sentences like they believe there to be several people missing and they believe the shit to be hitting the fan soon are perfectly grammatical tells us, in conjunction with the availability of wide scope for every defendant in (33), that ECM constructions with believe must support a derivation, in addition to the one involving ‘raising to object’, in which the ECM subject is not in the ‘object-of’ position (SpecRP2). That derivation is (41).

A derivation of this sort is not customarily posited for believe-type ECM constructions: the standard approach is to treat the infinitival complement of believe-type ECM constructions as a ‘bare’ TP, not as a null-headed CP. There are differences, to be sure, between want-type and believe-type ECM constructions (for instance, with respect to passivisation: while John is believed to be guilty is fine, *John is wanted to have a seat is not, contrasting sharply with they want John to have a seat). Those differences suggest that ‘V’ in the structures in (36) and (41) has a lexically variable ability to make the CP in its complement transparent to certain syntactic operations, including NP-movement: while V can open up its CP complement for NP-movement in (41), it cannot in (36). This lexical variability is closely tied in with the variability seen in the realm of ‘restructuring’. I have no substantive contribution to make to this thorny problem, which has been on the research agenda for several decades. But it seems to me that the differences between believe-type and want-type ECM constructions should not necessarily stand in the way of an analysis of the former in terms of CP-complementation, as in (41).

We now arrive at the conclusion that all infinitival ECM constructions of English can involve CP-complementation — not just the want-type cases but also the make out and believe-type cases. This may make it a little bit easier to account for the fact that English differs from many of the other Indo-European languages in having ECM with infinitival clauses: within Germanic, English differs markedly from Dutch and German in this regard; and the Romance languages, as a group, lack (straightforward) ECM with infinitival clauses as well.\footnote{In some of the Romance languages (French is one; see Kayne 1984), the problem with sentences such as *je crois Jean être intelligent ‘I believe Jean to be intelligent’ can be circumvented in a number of ways (via wh-movement, cliticisation or NP-movement of the subject of the infinitival clause). But without these additional operations, ECM with infinitival clauses consistently fails.} This is likely due to one simple lexical difference between these languages: while English has a case-licensing infinitival complementiser (with a zero allomorph), Dutch, German and the Romance languages do not. But note that this will only deliver a complete account of the difference between English and these other languages with respect to the distribution of ECM with infinitival clauses if (40) (with ‘raising to object’) is unavailable in the latter languages as well. The unavailability of (40) cannot be attributed to lexical quirks of infinitival complementisers: there is no C in (40). I see no solution for this yet.

Note that none of what I have said in the preceding paragraphs regarding the representation of ECM in constructions with propositional-attitude verbs affects the analysis previously proposed for resultative constructions. In particular, nothing I have said regarding the former impinges on the grammaticality of the structure in (7b), proposed for the ‘unspecified object’ reading of resultatives like he painted his hands black and blue. The important thing to bear in mind is that ‘OBJ’ in (7b) is not a proleptic object associated with the small clause in the V-complement position: there is no interpretive link at all between the unspecified object and the resultative small clause in (7b). In (37), by contrast, ‘OBJ’ needs to be a
place-hold er for the complement of V, which wants to get the ‘object-of’ interpretation. And precisely because proleptic objects cannot be associated with small clauses, (37) fails.

Before wrapping up, there is one more piece of business that needs to be addressed. In the structures of make out constructions proposed and defended in this paper, the particle spells out a head position (REL.2) in between v and V. How does this make it possible for v and V to ‘come together’ and be spelled out jointly as the verb make? I am taking a ‘late insertion’ approach to the lexicalisation of syntactic structures, following Distributed Morphology (Halle & Marantz 1993) and Nanosyntax (Starke 2009), and assume that the verb must spell out some portion of the Agree chain that v and V are both included in. The set of syntactic categories {T, REL.1=v, REL.2, V} in the structures under discussion in this paper together form a single Agree chain: (T, REL.1=v, REL.2, V). In English, the verb will be spelled out at v, never higher (there is no ‘V-to-T movement’ in English; no recourse to movement needs to be had in the approach taken here) and never lower (Modern English is not an OV language). In the verb-particle constructions that are the focus of the investigation here, a second element of the (T, REL.1=v, REL.2, V) Agree chain is spelled out as well, as the particle (at REL.2). Using the nanosyntactic notion of ‘spanning’ (see Svenonius 2012, Dékány 2011, also Brody 2000), we say that the verb (make in make out) spans T and REL.1=v, and the particle (out) spans REL.2 and V. Since make (qua span of the upper portion of the T/V-chain) and out (qua span of the lower portion of that chain) spell out different parts of a single chain (T, REL.1=v, REL.2, V), they in effect form a ‘complex verb’ — though make out is not represented as a V° unit, its ingredients (two spans involving different subparts of the chain) are part of a single Agree chain whose head is T and whose foot is V. The ingredients of the chain have relative syntactic autonomy. By taking a ‘late insertion’ approach that allows more than one part of the maximal T/V-chain to be realised individually, the proposal laid out here gives us the benefits of ‘complex verb’ approaches to verb-particle constructions of the make out type without having to countenance the more troublesome aspects of it (such as ‘excorporation’, or attachment of inflection inside a complex X°).

Once we allow the Agree chain (T, REL.1=v, REL.2, V) to have more than a single exponent, a range of other interesting possibilities presents itself. I will briefly mention two here. The first lies in the realm of so-called serial verb constructions. Serial verb constructions are clauses with multiple finite verbs sharing the same subject (and often the same object as well), all with the same tense value (see Stahlke 1970). The various verbal elements are all members of a single T/V Agree chain; take serials (such as (42a) from Yorùbá) are particularly interesting because in these, the take verb is arguably the representative of the ‘light verb’ v of the chain (spanning T and REL.1=v), and the second verb represents REL.2 and V, with the object sandwiched between the two in the SpecRP2 position (see (42b)). The take verb and the second verb are two individual exponents of the (T, REL.1=v, REL.2, V) chain, necessarily sharing the same tense value.

(42) a. mo mú íwé wá ilé
   I took book came house
   ‘I brought a book home’

   b. [REL.1=v=mú [REL.2 [V=wá [PP P ilé]]]]

The second possibility presents itself in the context of so-called semi-cleft constructions, found in several of the Romance languages (incl. European and Brazilian Portuguese, and Latin-American (though not Iberian) Spanish; see i.a. Wheeler 1982, Toribio 1992, 2002, Bosque 1999, Costa & Duarte 2001, Camacho 2006, Kato 1996, 2010, Mioto 2006a,b, 2008, Kato & Mioto 2012, Méndez Vallejo 2009, Resenes & Den Dikken 2012). The Portuguese sentence in (43b) is a variant of (43a) differing only in (a) the presence of a finite copular element between the verb and the direct object and (b) the obligatory contrastive focus reading of the object. The null hypothesis is that the same structure that underlies (43a) (viz., (44)) also serves as the source for the semi-cleft in (43b). The Agree chain (T, REL.1=v, REL.2, V) is realised singly in (43a), as the verb comprou ‘bought’ (i.e., comprou spans the entire set of heads in the Agree chain). Imagine now that the same chain has multiple exponents, with comprou spelled out in T and the finite copula foi spanning REL.1=v, REL.2 and V, and being spelled out in REL.1=v. Since comprou
and foi are exponent members of the same T/V-chain, they have the same tense and φ-feature inflections. This delivers exactly the output in (43b), with the copula immediately preceding the object and marking it for contrastive focus.  

\[(43)\]
\[
a. \text{ o João comprou um livro } \\
\text{ the João bought a book} \\
b. \text{ o João comprou foi um livro [‘semi-cleft’] } \\
\text{ the João bought was a book} \\
\]

If this maximally simple approach to semi-clefts is on the right track, it predicts that in secondary predication constructions, the copula should have more placement possibilities, because more positions in the T/V-chain immediately precede material that is contrastively focusable. And indeed, this is the case. Consider the examples in (45) and (46).

\[(45)\]
\[
a. \text{ o João pôs o livro na prateleira } \\
\text{ the João put the book on.the shelf } \\
b. \text{ o João pôs foi o livro na prateleira } \\
\text{ the João put was the book on.the shelf } \\
c. \text{ o João pôs o livro foi na prateleira } \\
\text{ the João put the book was on.the shelf} \\
\]

\[(46)\]
\[
a. \text{ o João considera a Maria inteligente } \\
\text{ the João considers the Maria intelligent } \\
b. \text{ o João considera é a Maria inteligente } \\
\text{ the João considers is the Maria intelligent } \\
c. \text{ o João considera a Maria é inteligente } \\
\text{ the João considers the Maria is intelligent } \\
\]

Following the logic of the preceding discussion, for the construction illustrated by (45) the derivation looks as in (47) (cf. (2), above), and (48) (cf. (38)) represents (46). We can again treat the semi-clefts as featuring T/V-chains with multiple exponents, the main verb spelled out in T and the additional copula spelled out either at REL1=V (immediately preceding the object) or at REL2 (preceding the secondary predicate).

\[(47)\]
\[
[TP \text{ o João}, [T \text{ [RP1=V} t, [REL1=V [RP2 o livro [REL2 [VP V [PP na prateleira]]]]]]]] \\
\]

\[(48)\]
\[
[TP \text{ o João}, [T \text{ [RP1=V} t, [REL1=V [RP2 a Maria] [REL2 [VP V [RP3 t inteligente]]]]]] \\
\]

Interestingly, there is an interpretive difference between (45b) and (46b) with respect to the size of the focused constituent. Whereas in (45b) contrastive focus, marked by foi, can be either on o livro or on the entire string o livro na prateleira, in (46b) contrastive focus is strictly confined to a Maria. The difference between the derivations for (45b) and (46b), based on (47) and (48), allows us to make sense of this. In (47), o livro is base-generated in the ‘object-of’ position, SpecRP2; in (48), on the other hand, a Maria...
has undergone ‘raising to object’, and is hence in a derived left-branch position. Derived left-branch positions are consistently narrow-focus positions — positions from which focus does not ‘project’ to the containing phrasal constituent.\(^\text{18}\) Thus, *JOHN studies linguistics* can be felicitously contrasted with *BILL studies linguistics*, but not with *BILL plays the guitar*. The minimal difference between (47) and (48) involving movement of the subject of secondary predication into the ‘object-of’ position is precisely what is responsible for the difference in information-structural properties of the semi-cLEFTs in (45b) and (46b).\(^\text{19}\)

Let me conclude. I have presented an approach to the syntax of the ‘object-of’ relation that represents it in terms of specification: the object function is projected onto the specifier of a RELATOR phrase projected immediately outside the minimal VP; whenever there is an external argument, the predication thus derived is in turn predicated of that argument, with the aid of a second RELATOR (now customarily referred to as \(v\)). The ‘object-of’ and ‘subject-of’ relations are structurally assimilated; the difference between them lies in the height in the structure at which they are established. Secondary predication constructions (including resultatives and constructions involving propositional-attitude verbs) played an important role in the discussion. We have been led to the conclusion that literal ‘raising to object’ genuinely exists: movement to the specifier position of the RELATOR phrase immediately outside VP is possible, and constitutes movement into a \(\theta\)-position. The frozen scope of sentences like *someone considers everyone smart* was derived from this, as was the fact that when focus is placed on the subject of the secondary predicate in the semi-cLEFTs of such constructions, we are always dealing with narrow focus. In the course of the discussion, I have unfolded an approach to some verb-particle constructions (in particular, those involving *make out, turn out and end up*) according to which the particle is outside VP, lexicalising the RELATOR-head involved in the establishment of the ‘object-of’ relation. This approach led me to rethink the link between V and \(v\) along ‘late insertion’ lines. Agree chains of multiple categories (in particular, T/V-chains) can have more than a single exponent, with each exponent lexicalising either a single position in the chain or a span of multiple heads. Such multiple exponence is found not just in verb-particle constructions featuring *make out, turn out and end up*, but also in serial verb constructions and semi-cLEFTs. The wider applicability of this idea beyond the cases discussed will, it is hoped, prove a fruitful area of future investigation.

References


---

\(^{18}\) This is doubtless related to the well-known fact that subextraction from derived left-branch positions is likewise impossible.

\(^{19}\) Needless to say, this simple sketch of the syntax of a few semi-cLEFT constructions does not constitute a complete analysis. See Resenes (2014) for much detailed discussion of the syntax of semi-cLEFTs in Brazilian Portuguese (not incorporating the approach to the spell-out of tensed-verbal chains outlined here).
Raising the Subject of the ‘Object-of’ Relation


FROM *ASPECTS*’ ‘DAUGHTERLESS MOTHERS’ (AKA DELTA NODES) TO *POP*’S ‘MOTHERLESS SETS’ (AKA NON-PROJECTION): A SELECTIVE HISTORY OF THE EVOLUTION OF SIMPLEST MERGE*

SAMUEL D. EPSTEIN  
The University of Michigan  
HISATSGU KITAHARA  
Keio University  
T. DANIEL SEELY  
Eastern Michigan University

1 Introduction

This squib briefly and very selectively explores the theory of human phrase structure generation and representation from *Aspects*\(^1\) to “Problems of Projection” (POP). Generally, our goal is to reveal the continuity of Chomsky’s always central goal: explanation via simplification. Our primary focus concerns labeling, projection, and the evolution of Simplest Merge in the quest for third factor reduction of Universal Grammar (UG).

Our central more specific point is that Merge, in its simplest form, \(\text{Merge} (X, Y) \to \{X, Y\}\), as recently developed in POP, eliminates both the mother node (i.e. projection) of standard Phrase Structure (PS) rules and representations, as well as the empty symbol delta (i.e. \(\Delta\)) postulated in *Aspects* (and in “A Minimalist Program for Linguistic Theory” (MPLT) as internal to Generalized Transformation (GT), see below). As a result, the POP system attains far greater depth of explanation, while, as Chomsky notes, concomitantly minimizing UG, thereby reducing the set of currently (and perhaps perennially) scientifically unanswerable questions regarding the evolution of UG.\(^2\)

---

\(^*\) For helpful discussion of certain of the ideas presented here, we thank Noam Chomsky and Chris Collins.  
\(^1\) We use the abbreviation *Aspects* to refer to Chomsky 1965.  
\(^2\) See Hauser et al. 2014.
2 A Brief Review of Certain “Aspects” of *Aspects*

Standard theory recursive PS rules of the sort postulated in *Aspects* (and earlier) provided a revolutionary solution to the long-standing paradox of discrete infinity: while the human brain is finite, the generative capacity of an I-language is infinite. Taking this as the core ‘creative’ property of human knowledge of syntax, it was Chomsky’s postulation of recursive PS rules that formally solved what had previously seemed paradoxical, namely “infinite use of finite means” (Humboldt’s term, as Chomsky notes). More generally Chomsky resurrected abandoned 17th century ideas regarding the mind, and centrally contributed to the modern day birth of the cognitive sciences, by noting the legitimacy of the postulation of abstract (non-material) concepts in normal (including of course, even physical) modern science, hence its permissibility, in fact necessity, in (previously, and still by many condemned) mentalistic theories of the brain.

Recursive PS rules provided an explicit representation of knowledge of linguistic structure – “the basic principle” in current terms – and accounted for the “creative aspect of language use.” A recursive structure-building mechanism is necessary for any adequate theory of I-language. But, given the commitment to explanation, important questions emerged regarding the phrase structure component of *Aspects*, some of which we are fully appreciating only now, some 50 years later. There are two questions we focus on here, involving the nature of the mother node (i.e. projection) and the nature of the empty symbol delta.\(^3\)

One central explanatory question that arises with *Aspect’s* conception of phrase structure and PS rules more generally is: “Why do we find that humans develop these particular (constructions-specific or category-specific) rules, and not any of an infinite number of other PS rules, or other types of rules?” Consider a PS rule like (1).

(1) \(\text{VP} \rightarrow \text{V NP}\)

Why exactly is there a VP on the left side of the arrow? Why more specifically is the mother labeled VP (and not, say, NP, or some other category or, for that matter, some non-category)? Is the label of the higher category stipulated, or does it follow from some deeper principle of generative systems, or perhaps from an even more general principle not specific to the human language faculty? We might even ask, Why is there a label at all? Indeed Collins (2002) proposed the elimination of labels, and Seely (2006) sought to deduce labelessness from the derivational theory of relations (Epstein et al. 1998, Epstein and Seely 2006). Since labels are not merged, they are in no relation, hence are equivalently absent from PS representations.

In *Aspects*, PS rules were essentially unconstrained, anything could be rewritten as anything and thus the existence and categorial status of mother labels were stipulated. Thus, for example, there was a “headless” rule like (2).\(^4\)

(2) \(\text{S} \rightarrow \text{NP VP}\)

---

\(^3\) Our primary concern is the cases where delta undergoes transformational substitution, as opposed to lexical insertion (see *Aspects* and below).

\(^4\) Note that there is an “unprojecting headed” rule like (i) in *Lectures on Government and Binding (LGB)*.

(i) \(\text{S} \rightarrow \text{NP INFL VP}\)

In this case too the mother is categorially unrelated to its daughters.
In (2), the mother node S is not a projection of (i.e. it is categorically unrelated to) its daughters. So, why do we find such headless phrases as S, while the major lexical categories seem to have heads, e.g., V in VP, and N in NP?

Another issue that arose involves a particular innovation of Aspects, namely, the postulation of the empty symbol delta $\Delta$:

“[S]uppose that (for uniformity of specification of transformational rules) we add the convention that in the categorical component, there is a rule $A \rightarrow \Delta$ for each lexical category $A$, where $\Delta$ is a fixed ‘dummy symbol.’ The rules of the categorical component will now generate Phrase-markers of strings consisting of various occurrences of $\Delta$ (marking the positions of lexical categories) and grammatical formatives.” (p. 122)

Thus, delta would appear in the (simplified) deep phrase marker associated with passive,

(3) $[S[\text{NP } \Delta] \text{ was arrested } [\text{NP the man}]]$

Raising of the object NP would involve the substitution of that object NP for the delta subject NP, yielding:

(4) $[S[\text{NP the man}] \text{ was arrested } [\text{NP (the man) }]]$

In effect, the delta is an empty and non-branching maximal projection; in this case an NP, and it has a purely formal status.

Note that both the mother (the S in (3)) and delta nodes of Aspects involve, at least in one sense, ‘look ahead;’ they are telic, anticipating transformation that will take place. And this in turn results, in part, from the ‘top down’ conception of phrase structure and from the particular model of grammar assumed in Aspects (in which (3) is postulated since $S \rightarrow \text{NP VP}$ and in which semantic interpretation is at the Deep Structure (DS) level (see Katz & Postal 1964)).

Aspects appealed to ‘top down’ PS rules of the sort in (1) and (2). But as pointed out by Chomsky (1995b), attributing the insight to Jan Koster, PS rules of this sort, where the mother node is projected, involve, at least in one sense, ‘look ahead.’ The mother node, the label VP of, say, $\text{VP } \rightarrow \text{V NP}$, is telic in the sense that it indicates the categories generated by the syntax that

---

5 According to Harris (1993), the delta node was developed from an earlier idea of Katz and Postal (1964).
6 It should be noted that the delta node played an important role in another crucial development in Aspects; namely, the separation of the lexicon from the syntactic component.
7 We are simplifying what is meant by ‘passive’ in this context. At the time passive was analyzed as involving two transformations—one moving the deep subject into the by-phrase and the other moving the object into the vacated subject position.
8 Chomsky (1995b: 318) states “Some of the general constraints introduced to reduce the richness of descriptive apparatus also had problematic aspects. An example is Emonds’s influential structure-preserving hypothesis (SPH) for substitution operations. As has been stressed particularly by Jan Koster, the SPH introduces an unwanted redundancy in that the target of movement is somehow ‘there’ before the operation takes place; that observation provides one motive for nonderivational theories that construct chains by computation of LF (or S-Structure) representations. The minimalist approach overcomes the redundancy by eliminating the SPH: with D-structure gone, it is unformulable, its consequences derived… by general properties of Merge and Attract/Move.” See below for further comment on the shift from Aspects to early minimalism.
will be relevant to the interpretive components, PF and LF. Put another way, the DS in (3) in fact encodes the categorial structure of what the Surface Structure (SS) will be. That is, if (structure preserving) delta substitution is required, then the NP subject of S is already present at DS, ’awaiting’ the obligatory arrival of the man. This encoding of SS in DS threatens the concept of level itself, suggesting that levels are in some sense intertwined, or non existent (as was later postulated in Chomsky 1993, Brody 1995, Chomsky 2000, Epstein et al. 1998, Uriagereka 1999).

Since in Aspects the initial level of representation, namely DS, fed into the semantic component, and since objects (presumably) required a label for semantic interpretation, it was necessary that labels be encoded as soon as possible, and hence encoded in the structures that served as input to semantics. Furthermore, since SS fed the Phonological Component, and since PF also (presumably) required labels, then the mother nodes needed to be present also at the last stage of a derivation.

Projections were needed, then, at both the initial (DS) and final (SS) levels, and were represented throughout the derivation. PS rules, projecting labels, – including NP delta projected from no lexical head N\(^9\), and the S-mother of NP delta, which is not a projection of NP delta, as in (3) above – served precisely the job of providing “sentential level” projection from the outset of the derivation (the theory being a theory of sentence-structure).

The delta is similarly ‘anticipatory.’ Recall a case like (3), in which a subject NP is already present structurally even before it is filled lexically, and this subject NP, since the mother S is already present, has no option of itself projecting; the mother S is already predetermined.

So, recursive PS rules of Aspects provided an empirically motivated, profound answer to a paradox solving the problem of discrete infinity. But, the nature of projection and of the empty symbol delta employed in Aspects raised a number of important questions, involving: projection, the nature of the lexicon, the relation between the lexicon and syntax, delta as a lexical item, the intertwining of levels, ‘look ahead’ the inviolability of S \(\rightarrow\) NP VP, and semantic interpretation at DS.

3 X-bar theory: the elimination of PS rules

X-bar theory represented a major development in the history of phrase structure.\(^{10}\) Rather than unconstrained, stipulated, and ultimately non-explanatory PS rules, the X-bar template imposed tight restrictions on what counts as ‘humanly possible phrase structure representation.’ X-bar theory sought to eliminate PS rules, leaving only the general X-bar-format as part of UG.

X-bar theory pushed endocentricity to its logical conclusion: since some categories seemed to be endocentric, like for example the lexical categories mentioned above (VP, NP, etc.), then, to eliminate unexplained asymmetries, it was assumed that all categories, including functional categories, are endocentric.\(^{11}\) Thus, a PS rule like (2), is excluded; it must be reduced to the X-

---

\(^9\) In Aspects, a rule \(A \rightarrow \Delta\), is proposed for each lexical category. Given this proposal, it follows that prior to lexical insertion, every lexical category, e.g. N, V, etc., is a delta. Thus, phrases like, say, NP are still headed, by N, in the sense that NP immediately dominates a delta N.


\(^{11}\) Of course, the asymmetry could have been resolved by alternatively assuming no category is endocentric. In current work the asymmetry is simply a consequence of recursive simplest Merge – with simplicity of the generative
Another major development within X-bar theory was that linear order was removed from the PS component; X-bar projections represented structure but not the linear order of elements within the structure. Standard PS rules simultaneously defined two relations, dominance and precedence, and therefore the application of a single PS rule could not (in retrospect) be a primitive operation since two relations, not one, are created. X-bar theory takes an important step in reducing the two relations to one, and it does so by eliminating linear order, which is a property of PF and (by hypothesis) not a property of LF. Thus, the theory came to express that “Syntax is not word order!” since word order is phonological. This disentangling of “dominance” (in hindsight, a misleading misnomer) and precedence, along with explaining their existence as subservient to the interfaces (dominance for semantics, precedence for phonology) was a profound step in the development of the standard strong minimalist thesis.\textsuperscript{12}

In X-bar theory, the mother is predetermined. Assuming binary branching, if X is non-maximal, its mother will be the category of X. If X is maximal, its mother will be the category of X’s sister. Consider the following tree representation (ignoring order)

\[(5) \quad \begin{array}{c}
XP \\
/ \ \\
YP \quad X' \\
/ \ \\
X \quad ZP
\end{array}\]

In (5), X and X’ are non-maximal\textsuperscript{13} and hence themselves project. YP is maximal and hence its mother is the category of YP’s sister (in this case X). The same holds for ZP. Projection from a head (i.e. endocentricity), and the syntactic representation of projection, are taken to be central concepts of X-bar theory, defining two core relations, (spec-head and head-complement).

What about the delta introduced in Aspects? It too implicitly remains in the X-bar schema. Under X-bar theory, the landing site of movement is often called SPEC, but SPEC is in effect a cover-term for delta as well. So, we could say delta was still assumed for movement under X-bar theory, i.e. X-bar was a constraint on transformationally derived structures in which projection is determined by X-bar schemata.\textsuperscript{14} So, the moving category has no chance to project — the mother of the mover ‘landing in’ SPEC is by definition not the mover.

X-bar theory thus raised a new set of questions – again, following the Chomskyan theme of a continued quest for yet deeper explanation. Namely, why should there be projection, and why should it conform to the X-bar template?

---

\textsuperscript{12} It should be noted that the removal of word order from syntax did not happen all at once. For example, the head parameter made explicit reference to word order. The 1980’s also saw the notion of directionality of government (Travis 1984). In Chomsky 2007, with the advancement of the primacy of CI hypothesis, however, it is clearly suggested that order is out of syntax; it is part of externalization. And we have an evolutionary hypothesis, consistent with the primacy of CI, i.e. language is primarily ‘for thought’ and not ‘for communication;’ see Chomsky (2013).

\textsuperscript{13} See Muysken 1982 on relational definitions of maximal and minimal categories.

\textsuperscript{14} Note that in this discussion we do not consider head movement nor adjunction (neither of which is movement to spec, and neither of which involves delta).
4 A Minimalist Program for Linguistic Theory: the Initial Transition from X-bar to Merge

Early minimalism brought major shifts in the architecture of the computational system for human language. In MPLT, Chomsky (1993) eliminates DS, and explicitly re-introduced delta into the theory, but it appears only internal to GT – not as a symbol appearing in/at a level of representation, as Aspects had delta in DS.\(^\text{15}\)

So, GT adds delta to a targeted category, a merging category substitutes delta, and the resulting structure is required to be consistent with X-bar schemata by definition. Thus:

\begin{quote}
“GT is a substitution operation. It targets K and substitutes K\(^1\) for Ø [delta] in K. But Ø is not drawn from the lexicon; therefore, it must have been inserted by GT itself. GT, then, targets K, adds Ø, and substitutes K\(^1\) for Ø, forming K*, which must satisfy X-bar theory … we never see Ø; it is subliminal, like the ‘first half’ of the raising of the NP to subject position.” (MPLT, p. 22)
\end{quote}

In the new theory, there are two distinct kinds of applications of GT. Binary GT takes two separate syntactic objects and combines them into a single object. Binary GT is thus the ‘ancestor’ of what would become External Merge. Singulary GT is the precursor of the more recent Internal Merge, where one of the objects being joined together is initially contained within the other. In effect, the elegantly constrained X-bar theory, together with its stipulated (or axiomatic) properties (endocentricity, head-complement, and spec-head relations) was taken to be a UG filter on transformational output representations.

5 Emergence: Bare Phrase Structure\(^\text{16}\)

X-bar theory represented a major step in the continued quest for explanation. But X-bar theory was not exempt from explanatory scrutiny. The question emerged: Why should X-bar theory hold? Why do we find these particular relations (endocentricity, head-to-complement, and spec-head), as opposed to an infinite number of alternative phrase structure systems? Stated in another way, and adhering to Minimalist method (see Chomsky 2007), we can ask: how “should” phrase structures be generated under minimalist assumptions? In Bare Phrase Structure, Chomsky’s (1995a: 396) answer was:

\begin{quote}
“Given the numeration N, C\text{HL} may select an item from N (reducing its index) or perform some permitted operation on the structure it has already formed. One such operation is necessary on conceptual grounds alone: an operation that forms larger units out of those
\end{quote}

\(^{15}\) It should be noted, however, that there is an interesting similarity between Aspects and MPLT with respect to the delta node: each application of Generalized Transformation taking X and Y, and forming Z may be regarded as a transformational mapping, and an empty dummy element delta exists only during this mapping.

\(^{16}\) The first part of the exposition of this subsection closely follows that of Epstein, Kitahara, Seely (2013), see also Epstein, Kitahara, Seely (to appear).
already constructed, call it Merge. Applied to two objects \( \alpha \) and \( \beta \), Merge forms the new object \( \gamma \). What is \( \gamma \)? \( \gamma \) must be constituted somehow from the two items \( \alpha \) and \( \beta \); ... The simplest object constructed from \( \alpha \) and \( \beta \) is the set \{\( \alpha \), \( \beta \)\}, so we take \( \gamma \) to be at least this set, where \( \alpha \) and \( \beta \) are constituents of \( \gamma \). Does that suffice? Output conditions dictate otherwise; thus verbal and nominal elements are interpreted differently at LF and behave differently in the phonological component ... \( \gamma \) must therefore at least (and we assume at most) be of the form \{\( \delta \), \{\( \alpha \), \( \beta \)\}\}, where \( \delta \) identifies the relevant properties of \( \gamma \), call \( \delta \) the label of \( \gamma \)."

Merge was introduced as the central structure building operation of the narrow syntax,\(^{17}\) necessary on conceptual grounds alone, and the simplest object \( \gamma \) constructed from \( \alpha \) and \( \beta \) by Merge was taken to be the set \{\( \alpha \), \( \beta \)\}. However, Chomsky (1995a, b) assumed the set \{\( \alpha \), \( \beta \)\} was too simple; it was assumed that empirical adequacy demanded some departure from the simplest assumption (the standard scientific tension between explanation and 'empirical coverage'); that is, the set must be labeled as in e.g. \{\( \delta \), \{\( \alpha \), \( \beta \)\}\}, where \( \delta \) identifies the relevant properties of \( \gamma \). Interestingly, note that in the above passage the argument for labels mentions only their necessity at the interfaces, and does not mention any reason for requiring them in the NS.\(^{18}\) We return to the status of labels in NS, momentarily.

So, the output of Merge is a labeled set \( \gamma = \{\delta \), \{\( \alpha \), \( \beta \)\}\}.\(^{19}\) In BPS, Chomsky (1995a:397-398) asked what exactly the label of \( \gamma \) is:

> “If constituents \( \alpha \), \( \beta \) of \( \gamma \) have been formed in the course of computation, one of the two must project, say \( \alpha \). At the LF interface, \( \gamma \) (if maximal) is interpreted as a phrase of the type \( \alpha \) (e.g. a nominal phrase if its head \( \kappa \) is nominal), and it behaves in the same manner in the course of computation. It is natural, then, to take the label of \( \gamma \) to be not \( \alpha \) itself but rather \( \kappa \), the head of the constituent that projects, a decision that also leads to technical simplification. Assuming so, we take \( \gamma = \{\kappa \), \{\( \alpha \), \( \beta \)\}\}, where \( \kappa \) is the head of \( \alpha \) and its label as well.”

Under this definition, the label of \( \gamma \) is the head of one of its immediate constituents; i.e. one of its members, assuming two-membered set theory. If \( \alpha \) projects, then the object \( \gamma \) constructed from \( \alpha \) and \( \beta \) by Merge is \{\( H(\alpha) \), \{\( \alpha \), \( \beta \)\}\}, where \( H(\alpha) \) is the head of \( \alpha \) (see also Chomsky 1995b). Additionally, the notion “term” (essentially the definition of “syntactic constituent”) is defined as follows: (i) \( K \) is a term of \( K \); and (ii) if \( L \) is a term of \( K \), then the members of the members of \( L \)

\(^{17}\) Ideally, there was a single operation but at the time there was appeal to both External Merge (X, Y are separate objects) and Internal Merge (Y contained in X). As we will see in some detail below, Chomsky (1995a,b) recognized that the simplest formulation of Merge is Merge (X, Y) \( \rightarrow \) \{X, Y\}, but he took it to be too simple, and proposed Merge (X, Y) \( \rightarrow \) \{Z, \{X, Y\}\}. Merge (External Merge) and Move (Internal Merge) are essentially the same. The crucial differences between Merge and Move are: Merge is a binary operation, while Move is singular operation, and the latter singulary operation is morphologically driven, and all sorts of conditions were assumed to constrain how it is morphologically driven, as we’ll see.

\(^{18}\) And, in fact, as argued in Seely (2006) since labels, as defined in Chomsky (1995a,b), are not syntactic objects, they are inaccessible to syntactic operations and are thus “syntaxically inert.” Labels in Chomsky (1995a,b) are not terms (as ‘term’ is defined there) and hence (informally speaking) ‘don’t exist.’ See Seely (2006) for detailed discussion.

\(^{19}\) For Chomsky (1995a,b), if (i) Merge is the only structure-building operation forming SOs, and (ii) each SO must be of the form \( \gamma = \{\delta \), \{\( \alpha \), \( \beta \)\}\} for the interface systems, then (iii) it must be Merge that creates a labeled set for the interface systems. We don't have anything other than Merge in NS.
are terms of $K$ (Chomsky 1995a:399).

Chomsky (1995a,b) did not discuss exactly how Merge operates to form such labeled sets, but he assumes that either $\alpha$ or $\beta$ may project (in principle), but if the wrong choice is made, deviance would result.\(^{20}\)

Notice, Chomsky (1995a,b) eliminated both delta and X-bar theory, but projection is still present; projection applies by definition. Merge was defined as $\text{Merge} (X, Y) \rightarrow \{Z, \{X, Y\}\}$, where $Z$ is either the head of $X$ or the head of $Y$. Under this definition, it was guaranteed that the label (= projected node) is either the head of $X$ or the head of $Y$, again by definition.

To sum up, (i) under phrase structure grammar with delta-substitution, a moving category has no chance to project by definition, (ii) under X-bar theory with SPEC/delta substitution, a moving category has no chance to project, again by definition, (iii) under GT with delta + X-bar theory (Chomsky 1993), a moving category has no chance to project by definition, (iv) under $\text{Merge} (X, Y) \rightarrow \{Z, \{X, Y\}\}$, where $Z$ is either the head of $X$ or the head of $Y$, either a hosting category or a moving category projects.\(^{21}\)

### 6 Simplest Merge: the Elimination of Labels and Projection from the Theory of Mental Representation

The strong minimalist thesis (SMT), presented by Chomsky (1993, 1995b) and elaborated by Chomsky (2000) and in subsequent work, takes the computational system for human language to be a “perfect system,” meeting the interface conditions in a way satisfying third factor principles.\(^{22}\) This is of course not an “assertion” but a hypothesis deemed worthy of exploration on a number of methodological grounds common to normal science.\(^{23}\)

Under SMT, therefore, the combinatorial operation of the generative procedure assumes (by hypothesis) the simplest formulation in what comes to be called “Simplest Merge”, a set-formation device that takes $X$ and $Y$, and forms $\{X, Y\}$.

(6) $\text{Merge} (X, Y) \rightarrow \{X, Y\}$

To the best of our knowledge, Collins (2002) was the first within the generative tradition to propose that labels be eliminated from the representation of syntactic objects and thus that the output of $\text{Merge} (X, Y)$ is $\{X, Y\}$ and not $\{Z, \{X, Y\}\}$. Taking Collins as the point of departure, Seely (2006) reanalyzes the matter derivationally, arguing that simplest merge (i.e. $\text{Merge} (x, y) = \{x, y\}$) is motivated on minimalist grounds alone and simplest merge entails the elimination of

---

\(^{20}\) We’re assuming that it’s deviance at CI, e.g. “chased the dog” labeled and hence interpreted as NP, rather than VP.

\(^{21}\) Certain previous analyses of mover-projection include projection of a moving (wh-phrase) maximal projection (Donati 2006) and the projection of a moving verb at LF (Epstein 1998). See also Hornstein and Uriagereka (2002).

\(^{22}\) See Epstein (2007) for detailed discussion of the idea that the theory is an “(internalist)-functional” one in the sense that the rules apply only “in order to” satisfy the interfaces. With the framework of Chomsky (2007, 2008, 2013, 2014) the idea is that operations freely apply as long as they conform to the laws of nature. There may be many functions, but there is no “the function.” Convergent derivations have the property of yielding interface interpretable representations that have been derived in a computationally efficient (meaning “laws-of-nature” compliant) way. In short, Merge freely applies as long as the applications conform to the laws of nature.

\(^{23}\) See Chomsky (2014) regarding standard misguided criticism that biological systems are ‘messy’—so it cannot be perfect.
syntactically represented labels and the elimination of (any type of) projection:

“It is important to stress that, viewed derivationally, it is not labels and projection that are eliminated in and of themselves, rather what is actually eliminated are two suboperations of the “complex” operation merge. It is a consequence of adopting the “simplest” version of Merge, namely, \([\text{Merge} (x, y) = \{x, y\}]\), that there are no phrasal labels nor projections, i.e. it is a consequence of the simplification of Merge that phrases are represented as in \([\{x, y\}]\), and not represented as in \([\{z, \{x, y\}\}]\). I’ll argue that this simplification of Merge is motivated on Minimalist grounds. The absence of labels is an immediate consequence of a well-motivated simplification of a fundamental, and arguably necessary, structure building (derivational) operation, namely Merge as in \([\text{Merge} (x, y) = \{x, y\}]\). In short, the question I am asking is: If indeed \([\{x, y\}]\) is the “right” type of representation, what is the nature of the generative procedure from which the relevant properties of these representations could be deduced?” (Seely 2006: 193)

Seely (2006) also argues that if Merge creates the only relations, then since labels (as in Chomsky 1995b) are in fact not merged, they are in no relation to anything; i.e. Seely seeks to deduce their absence from independently motivated proposals.

For Chomsky (2013), Merge, defined in the simplest form, also applies freely as long as it conforms to third factor principles (as it must) such as the proposed inclusiveness condition, “no new objects are added in the course of computation apart from arrangements of lexical properties” (Chomsky 1995b), and the no-tampering condition (NTC), “Merge of X and Y leaves the two SOs unchanged” (Chomsky 2008).\(^{24}\)

\(^{24}\) Given considerations of space, we can’t comprehensively consider a host of interesting questions regarding the (re-)development of ‘free’ application of rules, and just what ‘free’ means, and the extent to which there was a shift between constrained rules vs. free application (see Boeckx (2010) and Ott (2010) for important discussion). Given the SMT, in early minimalism there was computationally efficient satisfaction of the interfaces. The activity condition, Greed, Last Resort, and Criterial Freezing – (see Epstein 1992 in early MIN = Der constraints on Abar chain formation: once a wh phrase is in spec of +wh C, it is frozen – moving it would “undo” satisfaction of a filter seemed to be constraints on the application of rules. Free Merge seemed to represent a partial return to the ‘free’ application of move in GB. But, on one view, SMT itself hasn’t changed at all, in the following sense. In early minimalism, we eliminated all syntactic filters. Only naturalized interface filters, called bare output conditions, survived. As for syntactic constraints, we hope that they all reduce to principles of economy, now understood as third factor principles. So, what we had in early minimalism were: operations (e.g. Merge, Move), third factor principles, and bare output conditions. But operations in early minimalism were very complex. If you look at the definition of Move, for example, it has various sub-clauses beginning with “only if...” In other words, all the syntactic constraints on movement were “stipulated” as part of the defining properties of Move. But in subsequent work, there was some success in reducing those defining properties of Move to the third factor principles, and we now have the simplest formulation of Merge for both Merge and Move. Under POP+, what we have are: Merge (putting aside Agree), third factor principles (labeling by minimal search, NTC, inclusiveness), and bare output conditions. It is important to note POP+ adopts simplest Merge; the clauses beginning with “only if...” are all gone. In other words, there is no constraint on Merge (other than the laws of nature), meaning it is an operation that can apply freely. If so, the shift from early minimalism to POP+ is the success of reducing the additional defining properties of operations to third factor principles (and their interaction with Merge). When we had complex operations like Move, the system seemed deterministic, because the applications of Move are well constrained, but that was an illusion. It was always free to apply rules, but those rules themselves had necessary conditions as their defining properties, so the domain of free applications were limited. Say, if Merge is defined to operate on C and D only. Then, even if Merge is free, by definition, Merge cannot operate on M and D. Merge is defined to operate on C and D, but not on M and D, period. But this is part of the definition of Merge. So, it is still true Merge freely applies, but it cannot do anything beyond what it is designed to do. To sum up, when Merge/Move were defined with
In the absence of any stipulated constraints, there are just two applications of Merge. Suppose X is merged to Y (introducing the asymmetry only for expository purposes). Then, either X originates external to Y, call it External Merge; or X originates internal to Y, call it Internal Merge. Given NTC, Internal Merge yields two copies of X: one external to Y and the other within Y. There is no need to stipulate a rule of formation of copies (or remerge), and Chomsky’s (1993) copy theory of movement follows from “just IM applying in the optimal way, satisfying NTC” (Chomsky 2007). It would require stipulation to bar either type of application of Merge. Hence Chomsky explains the existence of what had been thought to be a very curious property, namely displacement and precedence; and shows that it follows from simplest unifying assumptions regarding the computational procedure (despite creating well-known parsing and communicative problems).

Notice, Merge does not encode a label.25 There is no labeled categorial node above X and Y present in the proposed NS mental representation. The categorial status of the set \{X, Y\} is representationally unidentified.26 That is, projection (or labeling) is no longer stipulated in the system as a defining property of Merge. Recall that standard PS rules simultaneously created dominance and precedence and so could not be primitive (as discussed above), likewise we could say that Merge with labels defining both mothers and sisters is not a possible primitive operation (though, in standard PS of course, sisters cannot be defined independent of mothers, nor can mother be defined independent of sister) Eliminating linear order, we create a set. This set is identified by minimal search, and at CI it then satisfies Full Interpretation (a bare output condition); a “higher node” or the “label” of an object formed by Merge is not represented in NS but is identified/computed by application of third factor minimal search — an (ideal) expectation under SMT.

Chomsky (2013) takes labeling to be the process of finding the relevant information of \{X, Y\} generated by Merge. He proposes that such labeling is “just minimal search, presumably appropriating a third factor principle, as in Agree and other operations.” So, labeling is not syntactically represented. No new categories are created in the course of a derivation (which, in fact, reduces to Inclusiveness). ‘Labeling’ is simply the name given to the independently motivated minimal search procedure, itself 3rd factor and hence not stipulated. POP then eliminates labels and projection, replacing it with a labeling algorithm that is an instance of the general principle of Minimal Computation, hence gaining yet greater depth of explanation.27

To understand how the labeling analysis, outlined in POP, works, let us examine the following two cases. Suppose SO = \{H, XP\}, H a head and XP not a head. Then minimal search will select H as the label, and the usual procedures of interpretation at the interface can proceed. By contrast, suppose SO = \{XP, YP\}, neither a head (recall PS rule (2) above). Here minimal

---

25 Note, if Merge does not encode a label, then label-identification cannot be a prerequisite for merge application within NS, as discussed in POP (see in particular p. 43, footnote 30).

26 See Collins (2002) and Seely (2006), among others, for earlier views of label-less syntax.

27 Note that eliminating labels from syntactic representations in fact comports with the standard view, in the following (counterintuitive) sense: (assuming binarity) we take an X and a Y and form one new object as a result of combining them. In current terms, this new object is the set containing X and Y and nothing else. Thus there is no ‘room’ for a fourth newly created object (i.e. a label) being created by combinatorial operations, again, assuming, as is standard that combinatorial operations create one and only one new object not present in the structural description.
search is ambiguous, finding both the head X of XP and the head Y of YP. This ambiguity is intolerable; left as is, labeling fails and Full Interpretation (FI) is violated at the interface levels. To resolve this so-called XP-YP problem for labeling, Chomsky (2013) suggests the following two strategies: (A) modify SO so that there is only one visible head, and (B) X and Y are identical in a relevant respect, providing the same label, which can be taken as the label of the SO.28

7 Conclusions: from Aspects to POP … and Beyond

The labeling analysis proposed in POP has many consequences (beyond those traced in POP itself), some of which are just beginning to be explored. Epstein et al. (2014; henceforth EKS), for example, argue that POP’s account of obligatory exit from intermediate positions in A'-movement, as in

(7) *you think \([_{CP} \text{ who } [_{Jon} \text{ likes }]]\)  \(\text{ (cf. who do you think Jon likes)}\)

carries over to A-movement as well. Thus, (7) is problematic at CI since the embedded CP will not be labeled. Similarly, EKS argue, for A-movement:

(8) *seems \([_{TP} \text{ a man } [_{to like Jon}]]\)  \(\text{ (cf. a man seems to like John)}\)

where the embedded TP fails to be labeled by CI. A host of technical mechanisms that have been proposed over the years to account for cases such as (8) are eliminated in favor of the single, simple labeling analysis. Chomsky (2014) extends his labeling-as-minimal-search analysis to the unification of the EPP and the ECP, reducing both to independently motivated properties of \(3^{\text{rd}}\) factor labeling. Chomsky (2014) also suggests an analysis of ‘obligatory halting’ in a Criterial Freezing position (see also Epstein, Kitahara and Seely to appear), reducing it too to labeling.

The 50 years since Aspects have brought many technical changes, but also ever deeper insight into the human language faculty. What in Aspects was two separate subsystems: (i) construction specific, category specific, language(s) specific and UG specific PS rules, accounting for word order and structure, and (ii) Transformations, accounting for ‘displacement,’ has now been reduced to a single, (maximally) simple, and necessary operation, namely, Merge \((X, Y) = \{X, Y\}\). This represents extraordinary progress in that the entire PS and Transformational components have been reduced to a single simple, linguistically unconstrained, operation, rendering explanation of the radically simplified theory of UG, conceivable and perhaps tractable so that the evolutionary ‘great leap forward’ of humans roughly 75 thousand

28 Note that in (B), the, or a, featural intersection of X and Y is taken as the label of \(\{XP, YP\}\). Interestingly, as pointed out to us by an anonymous reviewer of EKS (to appear), Chomsky (1995b) proposed this very idea, but at the time rejected the existence of projection of an intersect of features borne by X and Y, since X, Y might not have intersecting features. However, EKS (to appear) argue that the central phase-motivating (i)

(i) *there is likely \([_{XP} \text{ a man } [_{YP} \text{ to arrive}]]\)

is now deducibly disallowed precisely because there is no intersection of features between XP, YP and hence no label for (what would be) the embedded TP (with gibberish resulting at CI).
years ago, might one day be explained. And what were syntax-internal (and ultimately stipulative) constraints on rules or (non-explanatory) level ordered filters, are now by hypothesis being reduced to 3rd factor laws of nature. There have been many important discoveries since Aspects, and the current research program seems more intriguing and exciting than ever, particularly for those who are interested in scientific (theory-based) explanation and are willing to be surprised by that which seems so obvious, but demands (demanding) explanation.

References


ON THE ALLEGED ABSENCE OF CERTAIN AMBIGUITIES IN FREE CONSTITUENT ORDER LANGUAGES

Gisbert Fanselow
University of Potsdam

1 Introduction

Several grammatical phenomena in natural languages have so far not been amenable to a straightforward movement analysis within the limits of ‘standard’ generative grammar. Extraposition (Culicover & Rochemont 1990, Webelhuth 2013), scrambling (Bayer & Kornfilt 1994, Fanselow 2001), and head movement (Chomsky 2008) possess properties that are quite different from those that we find with classical cases of movement (leftward A- and A-bar movement). What these ‘difficult’ constructions seem to have in common is that they are (almost) exclusively concerned with the serialization of constituents. It is therefore tempting to assume that natural language grammars possess a separate serialization component, in addition to standard core syntax, with the rules of the former having properties quite different from the rules of the latter.

Such an idea may take various forms. Pesetsky (1998) argues for a grammatical model with a separate spellout component that has the architecture of an Optimality Theory account, while core syntax has a derivational (minimalist) shape. Rules of immediate dominance are also theoretically different from rules of linear precedence in grammatical frameworks such as GPSG and HPSG. The ingredients of c- and f-structure of LFG are very different, so that LFG also falls into the class of models that give serialization processes a special treatment.

An argument that serialization cannot be accounted for in terms of “normal” transformations can also be found in Aspects. Ever since I first read that argument in my 1977 introduction to linguistics class taught by Peter Staudacher, I have wanted to comment on the pertinent paragraph in Aspects, but for different reasons so in the course of the nearly four decades I am in linguistics now. At present, I think the data show quite nicely that the expressive power of syntax is rather limited – even though an additional linearization component is not called for.
2 Case Syncretism and Word Order

§4.4. of Chapter II of Chomsky (1965) discusses whether the strict connection between structural dominance and linear precedence that we find in standard syntactic trees and in standard syntactic rules is really adequate for natural language syntax. The chapter concludes with the claim that the free constituent order phenomenon has no real bearing on this issue. Chomsky arrives at this conclusion because free constituent order appears to have a property that “is not statable in terms of the theory of transformations” (Chomsky 1965: 127). The property in question is nowadays called ‘word order freezing.’

Consider in this respect German (1) that Chomsky uses to illustrate word order freezing with. The left-peripheral position of a German sentence (the “Vorfeld”/prefield) can be filled by either the subject (2) or the object (3) (among other choices) when the grammatical function of the phrase at the left edge is identified by morphological case (or by agreement). However, Chomsky states that when the morphology of the two noun phrases involved gives no unambiguous cue as to their grammatical functions, then only the subject can be placed into the prefield. In the absence of unambiguous morphology, word order freedom seems to disappear.

(1) Die Mutter sieht die Tochter
    the.nom/acc mother sees the.nom/acc daughter
    claimed to only mean: “the mother sees the daughter”

(2) Der Vater hört den Sohn
    the.nom father hears the.acc son

(3) Den Sohn hört der Vater
    the.acc son hears the.nom father
    “the father hears the son”

Being apparently unaware of the insights of Bierwisch (1963) concerning the nature of the placement of the finite verb in German, Chomsky interprets the data as follows: (2) represents the basic structure (SVO). There is a “stylistic inversion” rule of major constituents that can map (2) onto (3), but this inversion is inapplicable when “a structure” (one would nowadays say: a Phonetic Form) “is produced that might have been generated independently by the grammatical rules” (Chomsky 1965:127). Stylistic Inversion applied to (1) would not change the Phonetic Form of (1), and is therefore inapplicable to this sentence. For Chomsky, such a constraint on the applicability of the inversion transformation is, however, “not statable” in the theory of transformations, as mentioned above. Obviously, the constraint in question would have to be transderivational in nature (for each PF generated by movement M, one needs to check if could have also been built up without M), and is thus indeed incompatible with the theory of transformations developed in Aspects. Consequently, free constituent order cannot be the result of a transformational process if the data are properly interpreted.

Chomsky (1965:126) himself already notes a potential problem for his argument. When die Mutter “bears contrastive stress,” it can be interpreted both as a subject and as an object in (1), as exemplified by (4b,c)
On the Alleged Absence of Certain Ambiguities

(4) a. Der Vater liebt den Sohn, und die Mutter
   the.nom father loves the.acc son and the.nom/acc mother
   liebt die Tochter.
   loves the.nom/acc daughter

b. Den Sohn liebt der Vater, und die Tochter liebt die Mutter.
   “As for the son, the father loves him, and as for the daughter, the mother loves her.”

c. A: Den Sohn liebt die Mutter!
   the.acc son loves the.nom/acc mother
   “It is the son who the mother loves!”

   B: Nein! Die Tochter liebt die Mutter!
   No! the.nom/acc daughter loves the mother
   “No! It is the daughter who the mother loves.”

The presence of a particular contrastive stress on the phrase in the prefield is not a necessary condition for the acceptability of an object initial reading for (1), though. Die Tochter could also be a deaccented given object, and die Mutter an accented narrowly-focused subject, as in the constellation illustrated in (5)

(5) A: Wer liebt die Tochter?
   “Who loves the daughter?”

   B: Nun: Die Tochter liebt die Mutter!
   Well The daughter loves the mother
   “Well, the mother loves the daughter.”

So, the only constellation that seems to disallow an OVS interpretation for (1) is one in which both noun phrases are accented equally. But under such conditions, normally, an object cannot be placed before a subject even in a morphologically unambiguous sentence, cf. Fanselow & Levertova (2011) and Frey (2005) for different accounts of this fact. For an object to be able to precede a subject, there (nearly always) needs to be a difference in the information status of the two arguments. Without such a difference, it is not only the case that no object-initial reading is possible in (1), without such a difference, (3) is inacceptable, too. Thus, in a wide focus context, (6B) is not a felicituous utterance!

(6) A: What happened?

   B: #Einen Aufsatz hat ein Wissenschaftler veröffentlicht.
   an.acc article has a.nom scientist published
   “A scientist published an article.”

As an exception to this, objects may be fronted across subjects in wide focus contexts in order to be established as discourse topics, as in example (7) that could start a police report in the news. As the example shows, this works also in the case of a global morphological ambiguity.
Ein weinendes Baby rettete eine Passantin gestern aus einem verschlossenen PKW auf dem Parkplatz einer Vogelbeobachtungsstation. “A passerby rescued a crying baby from a locked car on the parking lot of a bird watching center.”

As soon as one realizes that object > subject order is always subject to pragmatic constraints, the alleged word order freezing effects disappear. Therefore, the judgments concerning (1) also do not motivate the postulation of an additional grammatical module consisting of rules of “stylistic inversion” and other rules of phonetic realization.

The purpose of these remarks is, of course, not to blame Aspects for not anticipating the outcome of a couple of decades of research on German word order. Rather, it is meant to illustrate that, very often, data that seems to have enigmatic properties under first inspection turns out to be amenable to a standard treatment after a thorough in-depth analysis.

3 Syncretism in Syntax, and More on Word Order Freezing

The alleged word order freezing effect explained away above involves Case syncretism: the syntactic distinction between nominative and accusative case is mirrored by a difference in morphological form only for masculine singular noun phrases. It is important to note that, while it does not affect reordering possibilities, the Case syncretism has syntactic effects, *viz.* in the domain of deletion/empty operators.

Many speakers of German respect a Case matching requirement in their formation of free relative clauses (cf., e.g., Vogel 2003) that renders only those structures acceptable in which the case assigned to the relative pronoun within the relative clause matches the Case assigned to the syntactic slot filled by the relative clause (8a,b). Morphophonological identity due to Case syncretism is sufficient for fulfilling the matching requirement: (8c) is grammatical, in spite of the fact that the free relative clause fills a nominative position in the matrix clause while the relative pronoun bears accusative case. That was represent both cases makes the structure acceptable.

\[(8)\]

<table>
<thead>
<tr>
<th>a.</th>
<th>Wer das mag liebt auch diesen Spinat.</th>
</tr>
</thead>
<tbody>
<tr>
<td>who.nom that likes loves also this.acc spinach</td>
<td></td>
</tr>
<tr>
<td>“whoever likes that will also love this spinach”</td>
<td></td>
</tr>
<tr>
<td>b.</td>
<td>*Wer das mag erfreut auch dieser Spinat.</td>
</tr>
<tr>
<td>who.nom that likes delights also this.nom spinach</td>
<td></td>
</tr>
<tr>
<td>“This spinach delights anyone who likes that.”</td>
<td></td>
</tr>
<tr>
<td>c.</td>
<td>Was du magst erfreut mich nicht.</td>
</tr>
<tr>
<td>what.nom/acc you.nom like delights me.acc not</td>
<td></td>
</tr>
</tbody>
</table>

Likewise, relative pronouns can be deleted in Bavarian dialects. As described by Bayer (1984), deletability is subject to a matching requirement similar to the one we saw for free relatives: the relative pronoun can be deleted only if its case matches the one of the head noun of the relative clause. But in the case of a syncretism, phonological identity of the cases suffices. Pullum & Zwicky (1986) list further cases where syncretism matters for syntactic processes,
On the Alleged Absence of Certain Ambiguities

... involving, e.g., processes of deletion in coordinate structures. So the data suggests that there is a set of syntactic rules that is sensitive to syncretism, and that is thus part of a “stylistic component” different from core syntax, but it is confined to rules of phonological realization (deletion, licensing of empty operators), much in line with Pesetsky’s (1998) suggestion.

We have seen that there is no word order freezing in cases of syncretism, but what happens in structures in which the two noun phrases bear syntactically identical cases? Indeed, the discussion of word order freezing in the East Asian languages has focused on such situations. Thus, it has been claimed (e.g. Flack 2007) that there is no reordering in Japanese double nominative constructions such as (9a), in which both arguments bear the same syntactic case.

(9) a. Hanako-ga Taroo-ga kowa-i
    Hanako-nom Taroo.nom afraid is
    “Hanako is afraid of Taroo.”

b. *Taroo-ga; Hanako-ga i kowa-i

But such Japanese examples function more of less just like the German example that Chomsky discussed. Flack observes (and tries to explain) that the scrambled structure (9b) becomes acceptable when Taroo-ga is focussed, but according to the judgments of Shin Ishihara (p.c.) structures such as (10a) improve when either argument is prosodically realized as an exhaustive focus, i.e., the structure improves when prosody marks an information structure difference between the two arguments. The same effect of licensing the inversion of the two arguments arises when particles such as –wa or –mo mark a difference in information structure.

(10) a. jishin-ga Taroo-ga t kowa-i
    earthquake Taroo-ga afraid is
    “Taroo is afraid of earthquakes.”

b. jishin-wa Taroo-ga t kowa-i
    earthquake-top Taroo-nom afraid is
    “As for earthquakes: Taroo is afraid of them.”

c. jishin-ga Taroo-mot kowa-i
    earthquake-nom Taroo-also afraid is
    “Taroo is also afraid of earthquakes.”

It is fair to conclude that it does not matter whether the identity of case form is due to case syncretism or the identity of syntactic case: Reordering of arguments is possible even when they are formally identical, as long as the two arguments differ in informational status. But the latter is simply the standard condition on reordering.

Similar constructions can also be found in German. In the rare cases of ECM in German, the subject and the object of the infinitival clause are both marked with accusative case. (11) shows that the accusative object can easily precede the accusative subject within the embedded clause.

(11) dass er das Baby weder dich noch mich t halten lassen wird
    that he the.acc baby neither you.acc nor me.acc hold let will
    “that he will let neither me nor you hold the baby”
Syntax, apparently, does not care whether the derived structures it produces involves a (global) morphological ambiguity or not when movement is at stake, but morphological ambiguity matters in the context of deletion. If there is a separate pronunciation component, it only affects the latter type of process.

4 A Ban on String-vacuous Scrambling?

In the paragraphs of Aspects under discussion, Chomsky not only argues for a separate stylistic component in grammar on the basis of so-called word order freezing, he also utilizes the phenomenon to illustrate the idea of a ban on deriving a syntactic object transformationally that might have also been generated ‘directly.’ While such a ban may be beyond the expressive power of core syntax in Aspects, the plausibility of “string-vacuous movement” has been discussed controversially in the last fifty years, and many syntactic models nowadays imply that movement should be avoided unless its application is forced by some more important grammatical constraint.

For free word order phenomena, it has been argued that there is a need for ruling out “string vacuous scrambling” (Haider 2003, Hoji 1995, Sabel 2005), which refers to a situation in which the original base order (12a) is restituted by a succession of several scrambling steps, as illustrated in (12c).

(12) a. dass der Mann den Adler beobachtete →
that the man the eagle watched
“that the man watched the eagle”
b. dass den Adler der Mann t beobachtete →
c. dass der Mann den Adler t t beobachtete

The need for such a ban on string-vacuous scrambling seems to stem from considerations of scope and binding. (13) and (14) represent the standard pattern of judgments for German (cf., e.g, Frey 1993). When subject and object are linearized in normal word order (13a), there is no scope ambiguity, while the marked object > subject order (13b) is ambiguous. Likewise, a pronoun can always be bound by a c-commanding noun phrase, irrespective of grammatical function (14a,b). The marked word order (14c) shows a reconstruction effect insofar as the pronoun contained in the object may be bound by the subject despite the lack of an overt c-command relation. In normal word order, a pronoun cannot be bound by a non-commanding argument, however (14d).

(13) a. dass ein Professor jeden Studenten kritisiert only: ∃ ∀
that a.nom professor every.acc student criticizes
“that a professor criticizes every student”
b. dass einen Studenten jeder Professor kritisiert ∃ ∀ and ∀
that a.acc student every.nom professor criticizes
“that every professor criticizes a student”
(14) a. dass jeder, seine Mutter liebt
that everyone.nom his mother loves
“that everyone loves his mother”
(14) exemplifies the janus-faced nature of scrambling very clearly: it creates new binding options (cf. 14b) and behaves like A-movement in this respect, but it also allows reconstruction (14c) and thus also functions like A-bar-movement.

Suppose, e.g., that scope is determined according to a rule such as (15), in the spirit of Aoun & Li (1993) and Frey (1993).

(15) \( \alpha \) can take scope over \( \beta \) if \( \alpha \) c-commands \( \beta \) or a trace of \( \beta \).

(15) forces us to exclude string vacuous scrambling. If (14d) also had the derivation (16), the object should be able to take scope over the subject, and bind the pronoun embedded in it. Therefore, it seems that we need to rule out a string vacuous series of applications of scrambling in order to predict that structures that appear in normal order are not scope ambiguous and show no “reconstruction” effects for binding.

(16) \( \text{seine}_i \text{Mutter jedeni} \text{liebt} \Rightarrow \text{jedeni}_i \text{seine}_i \text{Mutter} t_i \text{liebt} \Rightarrow \text{seine}_i \text{Mutter}_i \text{jeden}_t t_i \text{liebt} \)

This argument in favor of a ban on string vacuous scrambling faces a number of serious difficulties. First, as observed by Krifka (1998), scope inversion is possible in sentences with nominative > accusative order when they are pronounced with the fall-rise contour characteristic of sentences involving a contrastive topic. Why this contour should be able to lift the ban against string vacuous scrambling is fairly unclear, in particular, because it can be linked freely to any syntactic structure and thus cannot be considered a prosodic effect of scrambling (as Sabel 2005 has suggested).

Bans on string vacuous movement have often been discussed in the literature. The simplest case of a string vacuous movement can be easily dealt with in a derivational framework in terms of a statement affecting a single rule application: \( \alpha \) cannot move to \( \beta \) in \( \Gamma \) if the PF of the resulting structure \( \Gamma' \) is identical to the PF of \( \Gamma \). By such a formulation, (17a) would have to be chosen as a structure for a subject question rather than (17b).

(17) a. I wonder \([CP [TP who came]]\)
   b. I wonder \([CP who [TP t came]]\)

But the case of string vacuous scrambling is more complex. After all, the individual scrambling steps do change the word order of the syntactic object they apply to – it is only by their combination that the original order may be restored. A constraint that is meant to block such a constellation is necessarily transderivational in nature. For each structure \( \Delta \) with the sequence \( \ldots \alpha \beta \ldots \), with \( \alpha, \beta \) in derived positions, we need to check if there is a grammatical structure \( \Delta' \) which differs from \( \Delta \) only in that \( \alpha \), \( \beta \) occupy base positions. In that case, \( \Delta' \) would have to be assumed to block \( \Delta \).
It turns out to be extremely difficult to find an appropriate formulation for this ban on restoring a base-generated order. It cannot be a constraint that affects scrambling only, because the scope and binding pattern in (13)-(14) reappears in further structures. (18) is crucial in this respect. The object quantifier cannot bind into the subject in (18a), as expected, because it neither c-commands the subject in the resulting structure (18a) nor in any of the derivational steps that map (18b) through (18c) to (18a). But note there is a further derivation sketched in (18b-e-f-a) in which the object is first scrambled to the left of the subject (18e). (18e) makes binding possible, later, the subject is moved across the object to the prefield, in a movement step restoring the original relative order of the two arguments. This movement step should allow scope reconstruction, since it is an instance of A-bar-movement, but it does not. Notice that the final movement step has not been scrambling, but is rather movement to Spec,CP. Thus, a ban against sequences of scrambling that restore normal word order would not be sufficient to rule out a bound interpretation of the pronoun in (18a). The constraint in question must be more general.

(18) a. \( \text{seine}_1 \text{ Mutter liebe}\text{t jeden}_i \)  
   his mother loves everyone.acc
b. \( \text{seine}_1 \text{ Mutter jeden}_i \text{l}ib\text{et} \)
c. \( \text{lie}\text{b}\text{t seine}_1 \text{ Mutter jeden}_i \)
d. \( \text{seine}_1 \text{ Mutter t liebe}\text{t jeden}_i \)
e. \( \text{jeden}_i \text{ seine}_1 \text{ Mutter t liebe} \)
f. \( \text{lie}\text{b}\text{t jeden}_i \text{ seine}_1 \text{ Mutter} \)

But if we block any sequence of movement steps of A and B that, eventually, restores their original order, we would also rule out (19) (unless we give up the idea that subjects in Spec,TP originate in the VP): movement of the subject to Spec,TP places the subject to the left of the modal, but the modal then crosses the subject in a yes-no-question. In the interaction of head movement and A-movement, the restitution of the original order is grammatical.

(19) \( \text{can } [\text{vP the man [kiss Mary]]} \)  \( \rightarrow \)
   \( [\text{TP the man can } [\text{vP the man [kiss Mary]]} \)  \( \rightarrow \)
   Can [\text{TP the man can } [\text{vP the man [kiss Mary]]}]?

Can we circumvent this difficulty by restricting the necessary constraint to pairs of dislocations of arguments/adjuncts? Not really. When unstressed subject and object pronouns move to the left edge of TP in German to the so-called Wackernagel position, they retain their base order, as shown in (20). (21) illustrates that a combination of a movement to Spec,CP and a movement to the left edge of TP also may restore base order among the arguments.

(20) \( \text{dass er } \text{ is gesagt hat} \)
   that he it said has
   “that he has said it”
(21) \( \text{er hat's gesagt} \)

Furthermore, scrambling seems to be tied to certain grammatical phenomena that show that pairs of arguments must be able to scramble in such a way that they end up in the original
relative order. Thus, it has been argued (den Besten & Webelhuth 1990) that incomplete verb phrase fronting as in (22) is an instance of remnant movement: the arguments are scrambled out of the verb phrase before the latter is moved to Spec,CP. Note that the arguments evacuated from VP reappear in their base generated order in (22).

(22) \[ t_1t_2 \text{geküsst} \] dürfte wer_1 das Mädchen_2 haben
kissed might someone the girl have
“someone might have kissed the girl”

gelassen hat der Peter der Maria das Buch nicht
given has the.nom Peter the.dat Mary the.acc book not

If the licensing of parasitic gaps as in (23) involves the scrambling of arguments to the left of the adjunct infinitival (cf., e.g., Mahajan 1990), we again must conclude that scrambled phrases may surface in their original order.

(23) a. dass er den Studenten ohne vorher e gekannt zu haben einstellte
that he the student without before known to have hired
“that he hired the student without having known him”

b. dass er dem Kind das Buch anstatt e e zu geben weggenommen hatte
that he the.dat child the.acc book instead to give away-taken had
“that he had taken away the book from the child instead of giving it to him”

Crucially, as observed by, e.g., Fanselow (2001), all sentences in which the arguments \( \alpha \) and \( \beta \) appear in an order identical with their base linearization lack ambiguities of scope and reconstruction effects, irrespective of whether \( \alpha \) and \( \beta \) could still occupy the positions they were merged in (as may be the case for (13a)) or whether they have both undergone scrambling (as in, e.g. (24)).

(24) geküsst könnte jemand jeden haben only: \( \exists \forall \)
kissed could someone everyone have

The conclusion is inevitable that the absence of scope ambiguities in cases like (13a) and (24) cannot be linked to a ban on string-vacuous scrambling but must be due to some other constraints governing the interaction of PF and LF (cf., e.g., Bobaljik & Wurmbrand 2012, Fanselow 2012 for such proposals). But if scope and binding facts are unrelated to the presence of scrambling traces, the major reason for disallowing string vacuous scrambling sequences disappears! The complex derivation (18b-e-f-a) with ‘completely invisible’ scrambling can therefore be tolerated.

Finally, the argumentation against string vacuous scrambling suffers from the same difficulty that we have already discussed for the interaction of scrambling and morphological ambiguity: scrambling is not an operation that can be applied arbitrarily, rather, it serves various purposes, and it is hard to imagine that a fully string vacuous scrambling could fulfil any of these purpose. In (25), scrambling arguably applies in order to allow focal am Samstag to appear as much to the right as possible – the two arguments scramble around the high temporal adverb with changing their relative order. But what could be the reason for order-preserving scrambling in (26)? It has
no effect whatsoever on the placement of accented and deaccented material that governs much of German scrambling.

(25) \textit{dass das Kind den Vater am \textsc{Samstag} besuchte}

\begin{flushright}
\text{that the child the father at-the Saturday visited}
\end{flushright}

\text{“that the child visited the father on SATURDAY”}

(26) a. \textit{dass das Kind den Vater besuchte}
b. \textit{dass das Kind den Vater \textsc{t} \textsc{t} besuchte}

If the constraints governing scrambling apply directly to surface structure, and if they consist of ordering statements such as nom > acc, definite > indefinite, etc., they could not possibly distinguish between (26a) and (26b), but a general dispreference for movement would rule out (26b). In a derivational interpretation of the constraints, the two scrambling steps mapping (26a) on (26b) might be motivated by different constraints, so that C1 first triggers the movement of the object before the subject, and C2, the movement of the subject across the preposed object. E.g., the constraints animate > inanimate and definite > indefinite make different predictions for the linear arrangement in (27). Müller (2000) and Titov (2013) argue that the various constraints are ranked. Interpreted derivationally, this implies that a lower ranked constraints could not move a phrase if that implied the violation of a higher-ranked constraint. Only in the (presumably) rare case of two constraints having equal weight could we get into a potential loop of scrambling operations.

(27) \textit{dass der Stein ein Kind beunruhigt}

\begin{flushright}
\text{that the stone a child worries}
\end{flushright}

5 Conclusion

When I first read \textit{Aspects} in 1977, its interest seemed to be more or less historical in nature, having been replaced by Generative Semantics or Case Grammar on the one hand, and by Montague Grammar, on the other, and semantics seemed to be much more appealing to a beginning linguist than syntax. Still, the Pisa lectures were only 2 years away, starting a fascinating expansion of syntax research that would dramatically increase our knowledge of the limits of natural language grammar.

\textit{Aspects, so it seems to me now, was just the right preparation for this. It developed a very precise syntactic model that was quite constrained in nature – recall, e.g., the strict connection between serialization and hierarchy imposed by the phrase structure component, but recall also the quite restricted nature of the transformational component that did not allow, e.g., transderivational constraints to be employed. When such more complex tools appeared necessary in a certain domain, Chomsky proposed to isolate these more complicated regularities from syntax proper.}

Our discussion of free constituent order has shown that there is no need for transderivational accounts in syntax, or at least not in the domain that we considered. I take the situation with scrambling to be representative: when you look at a poorly analyzed data set, you may get the impression that it has odd properties. Closer inspection then reveals that there is an option for analyzing the data in a conservative fashion. Also, the data teaches us that it is always very
helpful to not simple formulate a putative constraint in prose, but to try and formulate it in every
detail.

Of course, one cannot show that this conclusion can be generalized, given the limited space
here. But other examples come to mind in which architecturally complex constraints on rule
application have evaporated, like the Minimal Link Condition that had, eventually, to be given
up. After all, the general strategy in Aspects for rule applicability was not so bad after all: a rule
can be applied whenever the phrase marker in question has a certain, locally determined
property. Full stop. No further fancy stuff required.

References

Bayer, Josef and Jaklin Kornfilt. 1994. Against scrambling as an instance of move alpha. In
Studies on scrambling, eds. Henk van Riemsdijk and Norbert Corver, 17-60. Berlin: de
Gruyter.
Bobaljik, Jonathan and Susi Wurmbrand. 2012. Word order and scope: transparent interfaces and
the \( \frac{3}{4} \) signature. Linguistic Inquiry 43: 371-421.
Freidin, Carlos P. Otero and Maria-Luisa Zubizarreta, 133-166. Cambridge, Mass.: MIT
Press.
Fanselow, Gisbert. 2001. Features, \( \theta \)-roles, and free constituent order. Linguistic Inquiry 32: 405-
437.
Fanselow, Gisbert and Denisa Lenertová. 2011. Left-peripheral focus: Mismatches between
Flack, Kathryn. 2007. Ambiguity avoidance as contrast preservation. In UMass Occasional
Papers in Linguistics 32, eds. Leah Bateman, Michael O’Keefe, Ehren Reilly and Adam
Werle. Amherst, MA: GLSA.
Gothoburgensis.
Haider, Hubert and Inger Rosengren. 2003. Scrambling: Untriggered chain formation in OV-
dissertation, University of Washington.


A NOTE ON WEAK VS. STRONG GENERATION IN HUMAN LANGUAGE*

NAOKI FUKUI
Sophia University

The last section of Chapter 1 of Aspects states that “[p]resumably, discussion of weak generative capacity marks only a very early and primitive stage of the study of generative grammar. Questions of real linguistic interest arise only when strong generative capacity (descriptive adequacy) and, more important, explanatory adequacy become the focus of discussion” (p. 61). This is a clear and explicit statement of Chomsky’s perception of the linguistic relevance of formal and mathematical investigations of grammars and languages, which he had initiated and explored in the 1950s and early 1960s. Clearly the concept of weak generation (the size of the set of output strings) plays only a marginal role, if any, in the formal investigation of human language. Rather, the nature of human language lies in its strong generative capacity, i.e., what kind of structural descriptions it generates, and more deeply, what kind of generative procedures it involves in generating those structural descriptions.¹

However, this understanding has been largely ignored and forgotten in the literature of formal investigations of grammars and languages, particularly in fields such as computational linguistics, biology, or more recently, neurosciences.² Thus, many recent studies in the brain science of language (or in the study of animal communications as compared to human language) deal with the alleged formal properties of languages, e.g., local dependencies, nested dependencies, cross-serial dependencies, etc. However, as indicated in many places in the literature (see note 2), these studies suffer from two major drawbacks.

One, the object “languages” are all finite, showing no discrete infinity. Since finite languages are outside of (or “below”) the popular hierarchy (the so-called Chomsky hierarchy) assumed in

---

*I thank Noam Chomsky for his comments on an earlier version of this paper. The research reported in this paper was supported in part by a Core Research for Evolutionary Science and Technology (CREST) grant from the Japan Science and Technology Agency (JST).

¹ See Chomsky (1986), Chapters 1 and 2 for much relevant discussion.

² References are too numerous to mention. See Everaert and Huybregts (2013) for an illuminating review. See also Ohta et al. (2013a, b) and the references cited therein for a detailed discussion on the importance of hierarchical structures and the Merge operation in the brain science of human language.
the relevant literature, the finitary status of the object languages is crucial, rendering the comparison between, say, finite-state properties and context-free properties simply irrelevant. This drawback can be partially overcome by some clever experimental tricks, such as letting the subjects “suppose” that the object language exhibits discrete infinity – even though it is, in fact, necessarily finite in actual experiments. The possibility of such a treatment indicates something significant about human cognition. That is, when objects proceed from one, two, and three, humans are most likely led to suppose that the sequence of that object will go on indefinitely, rather than stopping at some arbitrary number. This is certainly true for natural numbers, and should also hold when the object is a linguistic element. While this is an important point, I will not delve into this issue any further here.

Second, the focus of discussion so far in the literature has been on formal dependencies defined on the strings, and little attention has been paid to the abstract (hierarchical) structures assigned to the strings. This is a serious problem – because, as indicated ever since the earliest phase of contemporary generative grammar (see above), the nature of human language lies in the way it assigns abstract structures to strings (of words, phonemes, etc.) and not the set of such strings per se. That is, one should look for the nature of human language in its “strong generation” (the set of structures) – or ultimately, its procedures strongly generating the structures – rather than its “weak generation,” which is related to the set of strings it generates.

From this viewpoint, it is not at all surprising that human language does not quite fit nicely into the hierarchy in terms of weak generation (the Chomsky hierarchy). Human language is clearly beyond the bounds of finite-state grammar, and is “mildly” beyond the scope of context-free phrase structure grammar, but perhaps stays within the weak generative capacity of a certain subclass of context-sensitive phrase structure grammar (Joshi 1985). However, these results may not imply any substantive point, because even if a given grammar is adequate in its weak generation, the potential inadequacy in its strong generation will be sufficient for its exclusion as an appropriate model of human language. This is indeed true in the case of phrase structure grammars, regardless of whether it is context-free, where it seems inadequate even in its weak generation, or context-sensitive, where the way it strongly generates the structures for human language clearly exhibits its inadequacy. In short, the hierarchy in terms of weak generation, despite its naturalness and usefulness for other purposes, simply cannot provide a relevant scale along which human language is properly placed.

Let us consider some of the well-known types of dependencies in this light. For concreteness, let us take up the artificial languages discussed in Chomsky (1956, 1957): L₁ = {aⁿbⁿ} (n ≥ 0) (counter language), L₂ = xxⁿ (x ∈ {a, b}, xⁿ stands for the reversal of x) (mirror-image language), and L₃ = xx (x ∈ {a, b}) (copying language). All these languages are shown to be beyond the bounds of finite-state grammar, and L₁ and L₂ are shown to be within the bounds of context-free phrase structure grammar. L₃, on the other hand, is proven to be beyond the scope of context-free phrase structure grammar, since it shows cross-serial dependencies; however, its sentences can be generated by context-sensitive phrase structure grammar.

How does human language fit in this picture? It is well known that human language exhibits nested dependencies (cf. L₂) all over the place. It is also observed that human language sometimes shows cross-serial dependencies (cf. L₃). If we look at the actual cases, however, the distribution of these dependencies (among terminal elements) remains rather mysterious. For example, consider the following schematic Japanese example.

(1) NP₁-ga NP₂-ga NP₃-ga … NPₙ-ga … Vₙ-to …V₃-to V₂-to V₁ (-ga = Nom, to = that)
If we have \( n \) number of NP-gos and \( n \) number of Vs in this configuration, we can only have NP-V matching, as indicated above, i.e., the nested dependency. Thus, “John-ga Bill-ga Mary-ga waratta (laughed) to omotta (thought) to itta (said)” can only mean “John said that Bill thought that Mary laughed.” It can never mean, for example, that Mary said that Bill thought that John laughed (a case of cross-serial dependency). In a configuration such as (1), a typical sentence-embedding configuration, the linking pattern forming a nested dependency is the only possible one.

On the other hand, if we have the following coordinate configuration:

(2) NP\(_1\)-to NP\(_2\)-to NP\(_3\)-to ... NP\(_n\)-(to-)ga, (sorezore (respectively)) warai (laughed) (V\(_1\)), omoi (thought) (V\(_2\)), hanasi (spoke) (V\(_3\)), ..., itta (said) (V\(_n\))

[to = and; it can be replaced by other elements with the same meaning, e.g., ya (and)]

it can mean that NP\(_1\) laughed (V\(_1\)), NP\(_2\) thought (V\(_2\)), NP\(_3\) spoke (V\(_3\)), ..., and NP\(_n\) said (V\(_n\)). That is, a cross-serial dependency is certainly allowed here. In addition, other dependencies are also possible. Specifically, the so-called “group reading” is possible, where the interpretation is such that a group of people comprising NP\(_1\), NP\(_2\), ..., NP\(_n\) collectively laughed, thought, spoke, ..., and said. The nested dependency and other (mixed order) dependencies are probably impossible to obtain. The question is why this should be the case. Why is it that in a configuration like (1), only nested dependencies are allowed, whereas in (2), group readings and cross-serial dependencies (but perhaps no other dependencies) are allowed?

The answer to this question is difficult to obtain if we only look at terminal strings. However, if we look at the structures of these configurations, the answer seems obvious. In (1), neither of the sequences NP\(_1\), ..., NP\(_n\) (subjects of different clauses) and V\(_1\), ..., V\(_n\) (predicates belonging to different clauses) forms a constituent, whereas in (2), each of the sequences NP\(_1\), ..., NP\(_n\) (NP\(_1\) and NP\(_2\) and ... and NP\(_n\), conjoined NPs) and V\(_1\), ..., V\(_n\) (V\(_1\)-V\(_2\)- ... -V\(_n\), conjoined predicates) forms a constituent. Thus, the generalization here can be stated as follows: cross-serial dependencies in human language are possible only when the relevant terminal elements form a constituent. This constituency requirement strongly suggests that a transformation – a structure-dependent operation – plays an important role here. The actual way of deriving the cross-serial dependencies may be due to Copying Transformation, as suggested in Chomsky (1957) (p. 47, note 11, Chapter 5) – although in the cases considered here, “Copying” cannot be literal copying and should be characterized in a more abstract way, abstracting away from terminal elements and their immediate categorial status and focusing only on structural isomorphisms.\(^3\) Note that such a copying operation is more readily formulable in the current Merge-based system than in the classical theory of grammatical transformations. While grammatical transformations in the earlier theory operate on terminal strings with designated structures (structure indices), Merge directly operates on syntactic objects (structures), rather than strings, and with no direct reference to terminal strings which are not even “strings” at the point where Merge applies, since terminal elements are yet to be linearized. In this sense, Merge is even more structure-oriented than classical transformations, and leaves room for an operation that cannot be performed by classical transformations.\(^4\)

\(^3\) See Stabler (2004) for some related discussion on copying operations.

\(^4\) In various other respects, Merge is much more restricted and thus much less powerful than classical transformations (and phrase structure rules). Although there has been much interesting work in the literature (see, among others, Stabler (2010), Collins and Stabler (2011), and references cited therein), the issue of the overall
The observed linking patterns can be accounted for along the following lines. In (1), as we mentioned above, the nested dependency is naturally obtained by applying Merge in a phase-by-phase fashion, a conventional way of embedding sentences. Since neither the sequence of NPs nor that of predicates forms a constituent, the group reading is impossible. Cross-serial dependency is also impossible, because there is no structural basis for such a dependency. By contrast, the NPs and the Vs in (2) each form a constituent, and since there is a “copying” relation between the two constituents (the NP$_1$, …, NP$_n$ sequence and the V$_1$, …, V$_n$ sequence), group reading, which only requires the matching constituents (with no requirement on the order of terminal elements), is readily possible. Cross-serial dependency is also possible under the assumption that the structures generated by Merge ought to be maximally preserved through the linearization process, i.e., the structural properties should be directly mapped onto sequences, yielding a kind of maximal “copying” relation between the two constituents created by copying Merge. Nested dependency would force a departure from maximal copying, and hence is virtually impossible. Other mixed-order dependencies are even more difficult to obtain, yielding unintelligible interpretations.

If this line of reasoning is on the right track, the abovementioned generalization can be stated as follows:

(3) Cross-serial dependencies are possible only to the extent that they are “transformationally” derivable.

That is, item-by-item matching on terminal strings – which is necessary when the relevant item sequences do not form a constituent, as in case (1) – is not possible in human language. Consequently, context-sensitive phrase structure grammar is also disqualified in this regard as an adequate model for human language (cf. Chomsky (1963) for relevant discussion).

Given the basic properties of the current Merge-based system briefly discussed above, the same insight can be stated even in a more general (and stronger) form, which constitutes the main hypothesis of this paper.

(4) **Hypothesis**: Dependencies are possible in human language only when they are Merge-generable.

This is a generalization that cannot be made when the rule system of human language is divided into phrase structure rules and grammatical transformations, and when only a weak generation of human language is intensively investigated.

---

The generative power of the Merge-based system, particularly with respect to its strong generative capacity, seems to remain largely open. In the current Merge-based bare phrase structure theory, where no specification of linear order and labeling is made by Merge itself, the status of such classical concepts (and all the “theorems” and generalizations based on the concepts) as right- vs. left-branching structures, nesting vs. self-embedding structures, cross-serial dependencies, etc. does not seem obvious. Thus, Merge alone cannot of course distinguish right- vs. left-branching structures. The distinction should be made in the process of linearization (phonology). Merge, applying phase-by-phase, does generate nesting, as discussed in the text, but it underspecifies it, specifying no linear order. The distinction between nesting and self-embedding cannot be made by Merge, but it ought to be handled in the course of labeling. And so on. Also, even in narrow syntax, most of the dependencies are defined not (solely) by Merge, but via Agree, “predication,” and other miscellaneous relations/operations whose status is not crystal clear at this point. All of these problems should be cleared to seriously address the issue of the generative power of the Merge-based syntax.
Merge is a crucial operation (perhaps the only core operation) of the human language faculty, a biological endowment. Thus, when humans deal with a given dependency, the language faculty comes into play if the dependency is Merge-generable (directly or indirectly), but if the dependency is not Merge-generable, humans have to utilize other cognitive resources to handle it. More specifically, Merge does not specify linear order, does not count, and applies to structured objects (constituents) under certain structural conditions (characterizable by No-tampering, etc.). It follows, then, that $L_1$ (counter language) above cannot be dealt with by humans in terms of “finite-state grammar with counters,” since Merge does not provide counters. In this case, however, there is an alternative approach to characterize this language by Merge alone, without recourse to counters. Thus, this may well be the “right” account of $L_1$ if we are concerned with what is actually happening in the human brain. Note that in terms of weak generation, there is no right-or-wrong issue concerning the choice between the two ways to characterize $L_1$. $L_2$ (mirror-image language) is also Merge-generable. Thus, dependencies observed in (1) – nested dependencies – are easily generated by applying Merge cyclically (phase-by-phase). This is why nested dependencies abound in human language. However, cross-serial dependencies are Merge-generable only if the relevant items form constituents (note that Merge does not provide counters, nor does it give us linear order). Therefore, the distribution of cross-serial dependencies is rather limited in human language. Such a dependency is possible in (2), where the constituency requirement is fulfilled, but is disallowed in (1), where it is not.

Dependencies observed/defined on terminal strings are, in fact, epiphenomena, obtained as a consequence of Merge. Merge-generable dependencies are handled by the faculty of language, while non-Merge-generable dependencies are processed, perhaps as a kind of a puzzle or an intellectual exercise, by other cognitive capacities. It is thus predicted that the brain regions responsible for syntax, such as the left inferior frontal gyrus (L. IFG) and the left supramarginal gyrus (L. SMG) (Ohta et al. 2013a, b), will be significantly activated when Merge-generable dependencies are being processed, whereas brain activation patterns will be quite different when non-Merge-generable dependencies such as non-constituent cases of cross-serial dependencies are being processed. Reasonable and well-designed brain science experiments are likely to demonstrate these points. In fact, an experimental research is being conducted in an attempt to shed light on how Merge-generable and non-Merge-generable dependencies are differentiated in the brain (cf. Ohta et al. 2014).

In closing the discussion on generative capacity of grammars in Aspects, Chomsky argues for the utmost importance of explanatory adequacy/feasibility requirement as follows.

It is important to keep the requirements of explanatory adequacy and feasibility in mind when weak and strong generative capacities of theories are studied as mathematical questions. Thus one can construct hierarchies of grammatical theories in terms of weak and strong generative capacity, but it is important to bear in mind that these hierarchies do not necessarily correspond to what is probably the empirically most significant dimension of increasing power of linguistic theory. This dimension is presumably to be defined in terms of the scattering in value of grammars compatible with fixed data. Along this empirically significant dimension, we should like to accept the least “powerful” theory that is empirically adequate. It might conceivably turn out that this theory is extremely powerful (perhaps even universal, that is, equivalent in generative capacity to the theory of Turing machines) along the dimension of weak generative capacity, and even along the dimension of strong
generative capacity. It will not necessarily follow that it is very powerful (and hence to be discounted) in the dimension which is ultimately of real empirical significance. (p. 62)

These remarks, which appear to have been mostly forgotten or otherwise disregarded in the past fifty years or so, seem to hold true almost verbatim even now, or since the current theory is supposedly trying to go “beyond explanatory adequacy,” perhaps the empirically most significant dimension is to be defined by factors that lie beyond the feasibility requirement. The status of evaluation procedure that is behind the notion of “the scattering in value of grammars compatible with fixed data” in Aspects has been blurred particularly after the principles-and-parameters approach emerged around 1980.\(^5\) Thus, it is presently not even clear how to address what seems to be the most important theoretical question for the mathematical study of grammars. Perhaps it is premature to tackle such a question until we come up with a reasonable mathematical theory of the Merge-based generative system, which in turn should be based on a fuller understanding of the properties of human language. What has been suggested in this paper is a much more modest proposal, i.e., that we should shift our focus from weak generative capacity to, at least, strong generative capacity, i.e., the matter concerning descriptive adequacy, hoping for future development of the formal study of generative procedures themselves (Grothendieck’s theory of schemes comes to mind as a promising framework).

References


\(^5\) It is important to note in this connection that the logical possibility of eliminating the evaluation procedure is already hinted at (and disregarded) in *Aspects* (pp. 36-37).


A NOTE ON ENGLISH SUBJECT CONTACT RELATIVES

LILIANE HAEGEMAN
Ghent University/FWO

1 Syntax and Interpretation

Ever since the early days of generative grammar, syntacticians have been faced with the question as to the fit between syntax and interpretation. The following quotation (Chomsky 1965: 150) is as relevant today as it was in 1965:

It is clear from this fragmentary and inconclusive discussion that the interrelation of semantic and syntactic rules is by no means a settled issue, and that there is quite a range of possibilities that deserve serious exploration. The approach I have adopted in Chapter 2, § 3, is a conservative compromise between the attempt to incorporate the semantic rules strictly within the syntactic component and the attempt to elaborate the semantic component so that it takes over the function of the selectional rules.

Evidently, further insight into these questions will await a much more intensive study of semantic interpretive rules than it has yet been possible to undertake. The work of the last few years, I believe, has laid the groundwork for empirical investigation of this sort. There is a general theoretical framework parts of which have received empirical support. Within this framework it is possible to formulate certain reasonably clear questions, and it is also fairly clear what kind of empirical evidence would be relevant to deciding them. Alternative positions can be formulated, but for the present any one that is adopted must be extremely tentative.

In general, one should not expect to be able to delimit a large and complex domain before it has been thoroughly explored. A decision as to the boundary separating syntax and semantics (if there is one) is not a prerequisite for theoretical and descriptive study of syntactic and semantic rules.

---

1 Liliane Haegeman’s research is funded by FWO: 2009-Odysseus-Haegeman-G091409. Thanks to Lieven Danckaert, Marcel den Dikken, Alison Henry, Eric Lander, Terje Lohndal, Andrew Radford and Andrew Weir for helpful comments on an earlier version of this paper and thanks to them as well as to Alison Henry and Marcel Den Dikken for help with the judgments.
Over time, various positions have been held and this is not the place to provide an overview, but one trend that has come to the fore attempts to maximize the fit between syntax and semantics by “syntacticizing” the interpretive domains. An exponent of this approach is the cartographic line of work which, in the words of Cinque and Rizzi (Cinque and Rizzi 2010: 63):

The cartographic studies can be seen as an attempt to “syntacticize” as much as possible the interpretive domains, tracing back interpretive algorithms for such properties as argument structure . . . scope, and informational structure (the “criterial” approach defended in Rizzi 1997 and much related work) to the familiar ingredients uncovered and refined in half a century of formal syntax. To the extent to which these efforts are empirically supported, they may shed light not only on syntax proper, but also on the structure and functioning of the cognitive systems at the interface with the syntactic module.

In the wake of work in the cartographic tradition, specific analyses have sometimes been developed in which the syntactic formalization seems to be driven mainly or solely by interpretive properties. However, caution is required: while syntacticizing the interpretive domains the qualification ‘as much as possible’ should be taken seriously. I want to illustrate this particular point on the basis of the discussion two available analyses of so called subject contact relatives (SCR) in English, illustrated in (1), whose interpretive properties have led some researchers to postulate that what looks like a relative clause, the bracketed constituents in (1), is a root clause:

a. There was something [bothered me about the garage]. (Guardian, 23.3.1999, page 5, col 6)
c. She’d been told it was the stork [brought babies]. (Ruth Rendell, End in Tears, Arrow 2007, p. 169)

2 Relative clauses

One of the constants in the generative approach to syntax has been the assumption that filler gap dependencies can be modelled in terms of a movement derivation. In the first versions of the generative model, these derivations were formulated as ‘construction-specific’ transformations such as the passive transformation (Chomsky 1965: 23) and the relative transformation (Chomsky 1965:131). The original formulation of the transformations has changed over time and construction-specific transformations have given way to general instructions of the type ‘merge’ and ‘move’ in later versions of the generative enterprise, but the underlying intuition remains.

Leaving aside many differences of analyses and execution, in (2a) the relativized constituent originates as the object of arrested and moves to the left periphery of the clause.

a. The man whom they have arrested was not the thief.
In the simplified representations in (2b), the strikethrough represents the idea that the moved constituent originates as a complement of the lexical verb. The details of implementation are immaterial for the discussion at this point.

(2) b. \[\text{DP The man } [\text{CP whom } [\text{TP they have arrested whom}]] \text{ is not the thief.}\]

The restrictive relative clause is introduced by a relative pronoun, here whom; the relative clause is a subordinate clause and is taken to be part of a nominal constituent. The antecedent c-commands material within the relative clause, as shown, for instance, in (2c) which the negative antecedent no one licenses the NPI anything within the (subject) relative clause:

(2) c. There’s \[\text{DP no one } [\text{CP who can do anything about it}]].\]

Similarly, being part of a nominal constituent that is itself a constituent in a clause, the relative may also be in the scope of material in that containing clause: so, for instance, in (2d) the NPI anything in the relative clause is ultimately licensed by the c-commanding negative constituent (not/never) in the containing clause:

(2) d. There was never/not \[\text{DP anyone } [\text{CP who could do anything about it}]].\]

The pattern in (2a) alternates with two variants shown in (3): in (3a) the relative clause is introduced by the complementizer that, in (3b) there is no overt subordinating device, both relative pronoun and complementizer being absent. Examples such as (3b) are referred to as ‘contact relatives’. These examples can be brought in line with derivation (2b) if it is assumed that there is a null variant of the relative pronoun, represented as Ø in (4). In (4b) the complementizer is also non overt, a property which is independently available in other contexts in English, as shown in (5).

(3) a. \[\text{DP The man } [\text{CP that } [\text{TP they have arrested}]] \text{ is not the thief.}\]
b. \[\text{DP The man } [\text{CP they } [\text{TP have arrested}]] \text{ is not the thief.}\]

(4) a. \[\text{DP The man } [\text{CP Ø that } [\text{TP they have arrested}]] \text{ is not the thief.}\]
b. \[\text{DP The man } [\text{CP Ø Ø } [\text{TP they have arrested Ø}]] \text{ is not the thief.}\]

(5) I thought \[\text{CP (that) you were going to take the garbage out}].\]

The presentations above are obviously much simplified and the discussion does not do justice to all the work that has gone into formalizing English relativization, but it will do as a starting point for the discussion.

There is a consensus that in the unmarked case subject relatives with the null variant of the relative pronoun and no overt complementizer, or ‘subject contact relatives’ (SCR), a term first used by Jespersen (1961), are not acceptable:

(6) a. I asked one of my uncles who was an engineer and he told me...
b. *?I asked one of my uncles was an engineer and he told me... (Lambrecht 1988: 321, his (13b))
Yet, as shown by (1), SCR are attested and judged acceptable by some speakers: the following are just a sample of anecdotally collected examples from written sources; the examples all typically represent spoken English. Lambrecht (1988) provides additional attested examples from spoken English. Typical contexts in which the SCR is found are (i) presentational there sentences (7), (ii) presentational clefts (8), (iii) complements of have (9), (iv) predicates (10).

(7) a. He didn't seem to be in his flat either and there was something bothered me about the garage, which Damien normally propped open with a stick because the door was broken. (Guardian, 23.3.1999, page 5, col 6)
b. But there’s always something happens that’s new or different. (Guardian, G2, 18.4.2, page 4, col 3)
c. There was one driver stopped. (Ruth Rendell, End in Tears, Arrow 2007, p. 235)
d. There was one girl said she'd take Charlie out for free. (Ruth Rendell, The Keys to the Street, Hutchinson 1996, Arrow 1997, p. 219)
e. There's only one bugger in the whole city can make out what she's saying, and I've just packed him off home. (Ian Rankin, The Hanging Garden, Orion 1998, p. 45)
f. I'm ever surprised when I get cut that there's not a pile of dust runs out instead of blood. (David Storey, Pasmore, Longman 1972, Penguin 1976, p. 85)
g. Wherever there's drugs, there's money needs laundering. (Ian Rankin, The Hanging Garden, Orion 1998, p. 326)
h. There's days go by, weeks maybe, when I never see the news. (Bernard Mac Laverty, Grace Notes, Vintage, 1998, p. 15)
i. There’s a lot of clubs would give a manager a 10-year contract for doing that. (Observer 18.5.8 page 30, col 3)

b. He had some woman would visit him... red hair. (Ian Rankin, The Hanging Garden, Orion 1998, p. 189)
c. I have a woman comes in every day. (Bernard Mac Laverty, Grace Notes, Vintage, 1998, p. 113)

(9) a. When she was a child she’d been told it was the stork brought babies. (Ruth Rendell, End in Tears, Arrow 2007, p. 169)
b. It must be him overheard me talking in here. (Ruth Rendell, The Keys to the Street, 1996 Hutchinson, Arrow books 1997, p. 176)
c. Maybe, but was you told me that any departure from the norm is important. (Ruth Rendell, A Sleeping Life, An Inspector Wexford Mystery, 1964, Arrow 1994, p. 78)
d. What was it put Macbeth in mind? (Nicholas Freeling, Lady Macbeth, André Deutsch, 1988, Penguin 1989, p. 35)
e. What is it brings you here? (Blackadder Season 2 Episode 1 "Bells", (time: 4.58) example thanks to Lieven Danckaert, p.c.)
you were the one came in and told us that you'd taken a picture for a woman who was making a claim to criminal injuries. (Frances Fyfield, *Blind Date*, Corgi 1999, p. 65-66)

The contexts illustrated above are those usually discussed in the literature. Focusing on Hiberno English, Doherty (1993) also mentions complements of *know*:

(11) I know a smart Greek owns maybe twenty restaurants (W19)
    I know a fella can get all the tobacco he wants. Frank Dooley ...(F 102)

Doherty (1993, 1994) and Henry (1995) discuss SCR with special reference to speakers of Hiberno English, but as the sources of the attested data show, they are certainly not the only ones admitting the pattern. Quirk et al (1985: 1250) label SCRs in the contexts illustrated in (7) and (9) as ‘very colloquial’ and ‘less acceptable’ than the alternatives introduced by *that* or a relative pronoun. Similarly, Huddleston and Pullum (2002: 1055) signal similar examples and qualify them as falling ‘at the boundary between very informal and non-standard’.

What is common to all the examples of SCR illustrated above is their discourse function: what superficially is a matrix clause serves to introduce a focus, a nominal constituent which I will refer to as the ‘antecedent’ of the SCR; the SCR itself provides information on that focus and is the more informative part of the utterance. The discourse interpretive properties of such contact relatives were first highlighted in Prince (1981:247) and are discussed extensively in Lambrecht (1988).

Huddleston and Pullum (2002: 1055, their (68iii)) cite (12a) with the SCR modifying a subject nominal as ‘non-standard’:

(12) a. Anyone wants this can have it.

However, Doherty (1993) points out that while for many of the Hiberno English speakers he consulted, the SCR is restricted to the contexts in (7)-(11), more liberal speakers allow such patterns: (12b) and (12c) are attested. He concludes that some speakers have a more liberal use of SCR. The attested (12d) would probably fall under a more liberal variety.\(^2\)

(12) b. Anyone can help afterwards is welcome. (Doherty 1993: 107, his (78a))
    c. Everybody lives in the mountains has an accent. (Doherty 1993: his (79a))
    d. If we were married and you helped yourself to a piece, I'd throw the first thing came to hand at you. (*Independent on Sunday*, ABC, ‘Heart Shaped Box’, Joe Hill, page 24, col 1)

\(^2\) The precise distribution and the dialect appertenance of these patterns remains unclear but it is not relevant for the point to be made here.

\(^3\) Den Dikken (2005: 700) signals that SCRs in Appalachian English have a different discourse function. Since I don’t have access to these speakers, this variety is not discussed here.
3 The Syntax of Subject Contact Relatives

Accounts for the SCR have usually focussed on sentences such as those in in (7)-(10) above. They fall into two broad categories: some treat SCR as a variant of relative clauses, others treat them as a distinct phenomenon in which what looks like a relative clause is in fact a root clause.

3.1 The Subordinate Account

The relative clause account of SCRs is represented by Jespersen (1961) and in the generative literature by Doherty (1993: 111, 1994). For Doherty ‘[N]oun phrases modified by subject contact clauses ... have the external syntax of noun phrases’ (Doherty 1994: 58). To account for their particular properties, Doherty assumes SCRs lack the CP layer (see also Weisler 1980). One piece of evidence advanced by Doherty for the reduced structure is the fact that, unlike relatives with an overt pronoun and like object contact relatives, SCRs are incompatible with left peripheral adjuncts (the judgement is shared by Huddleston and Pullum 2002: 1055, on the different judgement in Harris and Vincent 1980, see Doherty 1993b: 62, note 4))

(13) a. The man *(who) years ago Mary used to know well. (Doherty 1993: 62, his (70))
b. That’s the girl *(who) just yesterday was talking about you. (Doherty 1993: 62, his (73))

I will not go into the details of Doherty’s analysis here (see Henry (1995) for an evaluation), but what is important is that for Doherty the SCR is a subordinate clause, it can constitute a unit with the nominal to its left and –depending on its precise function- it can be in the c-command domain of what would be matrix material (2c,d).

3.2 The Root Account

A radically different representation in the traditional literature (Erdmann 1980: 161), first developed in the generative literature by Henry (1995) and elaborated in Den Dikken (2005) closely matches the syntactic representation of the SCRs with their specific discourse function, with as a key property the observation that the SCR seems to carry the main assertion. Hence, it is proposed that what looks like a relative clause is in fact the matrix clause. In line with cartographic proposals in which discourse functional interpretive effects are encoded in left-peripheral functional projections (along the lines of Rizzi 1997 and much work after him), Den Dikken (2005: 698, his (14)) assigns to SCRs the representation in (14):

---

4 Andrew Radford finds the adjuncts not degraded. This is an important point that merits further study, but as it ties in with the general question as to the status of left peripheral adjuncts, it goes well beyond the scope of the present paper.
There's one woman in our street
Ø went to Spain last year

In other words, both for Henry (1995) and for Den Dikken (2005) –henceforth abbreviated as DD - the ‘matrix’ clause of the SCR (S1, *there’s one woman in our street*) is the topic of an articulated Topic-comment structure and the SCR (S2, *went to Spain last year*) is the root clause expressing the comment. The discourse function of the topicalized matrix clause is that of introducing the focus, the ‘antecedent’ of the SCR, which ‘will serve as the anchor for the comment clause’ (DD 2005: 698). Syntactically, S1 is compared to the Hanging topic illustrated in (15), whose function is also to introduce a nominal, *syntax*, on which the root clause provides a comment:

(15) a. As for syntax, many students find it too difficult a topic to pursue for their dissertation. (based on DD 2005: 703)

The interest of the Henry/DD topic-comment representation is obvious. First, the representation directly encodes or “syntactizes” in Cinque and Rizzi’s (2010) wording, the discourse function of the SCR: in the representation in (14), the SCR supplies the main assertion. See also Lambrecht (1988) for a similar approach in a Construction Grammar framework in terms of syntactic amalgams.

In addition, if the SCR is indeed a root clause, as in (14), then, as pointed out by Henry (1995) and by DD (2005: 698), it is not unexpected that it can feature a null subject, since many varieties of spoken English allow for null subjects in root domains (see Schmerling 1973, Trasher 1977, Napoli 1982, Rizzi 1995, 1999, 2006, Haegeman 2013, Weir 2009, 2012). The restriction on left peripheral adjuncts pointed out by Doherty (1993, 1994) would also follow since, as shown in Weir 2009, 2012, in spoken English (though not in diary style writing, see Haegeman 2013) subject omission is indeed incompatible with left peripheral adjuncts. Moreover, if the topicalized S1 in (14) is an instantiation of a Hanging topic as suggested by Henry (1995: 135) and by DD (2005: 698, note 6), its compatibility with the null subject is unproblematic because Hanging topics (15b) – as opposed to regular topics Haegeman 2013) (15c) - are compatible with root null subjects.

(15) b. As for syntax, Ø found it too difficult c. *Syntax, Ø found too difficult.

However, assimilating the null subject in SCR to the same phenomenon that underlies subject omission in spoken English is not straightforward. The distribution of null subjects in matrix

---

5 Thanks to Andrew Weir for pointing this out.
clauses and in SCRs is not the same. In particular, as shown by Napoli (1982) and Weir (2009, 2012), in spoken English cliticizable auxiliaries are unacceptable with subject omission but at least some of these are relatively acceptable in SCRs. So for instance, informal spoken English would not allow null subjects for (15d) and (15f) with the cliticizable auxiliaries will (cf. ‘ll) and has (cf. ‘s), but the corresponding SCR with these auxiliaries is acceptable (15e, 15g). With respect to SCR with actually reduced auxiliaries there is conflicting evidence, DD (2005: 659) rules it out⁶, but Lambrecht does provide the attested (15h).

(15)  
d. I’ve recruited some people. ?* Ø Will do the experiment for us.
e. I’ve recruited some people will do the experiment for us.
f. I met someone. ?* Ø Has been stopped by the police seven times.
g. ?I met someone has been stopped by the police seven times.
h. I’ve got a friend from Chicago’s gonna meet me downstairs. (Lambrecht 1988: 318, his (7))

4. A topic-comment analysis of SCR

In this section I examine the predictions by Henry’s and DD’s topic-comment representation in (14) of SCR and I will show that the representation raises problems both with respect to the internal relation between the components of the SCR pattern (section 4.1.) as well was for its distribution (section 4.2.).

4.1 Internal Syntax

4.1.1 Constituency

In Doherty’s (1993, 1994) account, SCR are relative clauses and may form a constituent with their antecedent. According to the topic-comment representation (14), however, the SCR clause does not form a constituent with the nominal expression which it provides information on (DD 2005: 702). This leads to the prediction that, not being a constituent, the ‘antecedent’ and the SCR will not be able to enter into coordination with another constituent. However, a string composed of such an antecedent and a SCR can be coordinated with a string composed of an antecedent and a regular relative clause, suggesting that the SCR does form a constituent with the antecedent. Such data favour an analysis in line with Doherty’s in which the antecedent and the SCR can constitute one unit of structure.⁷

(16)  
a. I have [one colleague runs a sushi shop] and [another one who has a burger restaurant].  
b. There’s [one student lives in a hotel] and [another one who lives in a renovated railway station].

---

⁶ Andrew Weir finds (15h) unacceptable.
⁷ Examples such as (12b) and (12c), in which the SCR modifies a subject nominal, are also problematic for the topic-comment analysis. However, these are not accepted by all speakers and are not necessarily intended to be covered by the topic-comment representation.
4.1.2 C-command

Given representation (14), constituents internal to S1 should not be able to c-command those in S2, the SCR. As shown by (17) this prediction is incorrect. In (17), the quantificational subject of S1 binds a pronoun in S2. In the attested (7i) above, too, the ‘antecedent’ a lot of clubs binds the nominal a manager, the latter thus has a variable reading. Also, a constituent in S1 can license an NPI in S2 (18).

(17)  
   a. Each of themi took a book belonged to the otheri.
   b. Every studenti is looking for a teacher speaks hisi language.
   c. Nobodyi took anything didn't belong to themi.

(18)  
   a. There’s no one can do anything about it.
   b. This was not something was ever considered in the discussion.
   c. She is not someone has ever been considered for a tenured position.
   d. It isn't something was ever considered in the negotiations
   e. That was the stormiest night was ever in this parish. (Doherty 1994: 58, his (40), LN12)

DD’s (14) is close to a paratactic configuration (see discussion in DD 2006, and also Gärtner 2001a,b and Ebert, Endriss and Gärtner 2007 on German embedded V2). But while SCR allow for NPI licensing out of S1, this is not possible in genuine parataxis, as shown in (19). The unacceptability of the NPI ever in the second conjunct in (19) can be attributed to a lack of c-command in the paratactic configuration.

(19) * They were not shortlisted and were ever considered for the position.

If lack of c-command is responsible for the ungrammaticality of (19), then the availability of bound pronouns and NPI licensing in SCR suggests that the relevant constituents in S1 c-command S2, casting doubt on the topic-comment representation (14). Note that these phenomena would be in line with Doherty’s relative clause analysis of SCR since it is expected that the relative clause is within the c-command domain of the antecedent and –depending on the position of the relativized nominal- may also be in the scope of other constituents of the containing clause (cf. (2c), (2d)).

4.2 External syntax

4.2.1 Main clause phenomena

For Doherty (1993, 1994), SCR have the distribution of relative clauses and are part of nominal constituents (1994: 58). Thus SCR are subordinate clauses. Henry (1995) and DD (2005) treat

---

8 Thanks to Andrew Weir for generous help with the data. Needless to say he is not responsible for the way I have used his comments.
SCR as topic-comment structures with root status. In particular, they compare the position of S1 in (14) to that of a Hanging topic and DD (2005:703) explicitly qualifies the pattern as unembeddable, appealing to the analogy with unembeddable Hanging topics in (20).

(20) *It’s unfortunate that, as for syntax, many students find it too difficult a topic to pursue for their dissertation. (Den Dikken 2005: 703)9

SCR are thus expected to be a root phenomenon in the sense of Emonds (1970), Hooper and Thompson (1973). However, this prediction is not borne out. SCR is compatible with domains that resist root phenomena, such as temporal clauses (21a), conditional clauses (21b) and complement clauses of factive verbs (21c). These are known to be incompatible with root phenomena (Emonds 1970, Hooper and Thompson 1973, Haegeman 2012).

(21) a. My head of department is reluctant to intervene for his male students but [whenever there's a woman wants to see him], he immediately will act.
   b. I'm not available but [if there's a blonde girl with glasses wants to see me], give me a call.
   c. I resent [that there are always so many students want to see me].

Observe that the fact that SCR may be embedded in adverbial clauses also means that the null subject postulated for S2 in the topic-comment account is unaccounted for, since English does not allow null subjects in these embedded domains: (22) is from Henry 1995: 128, her (43).

(22) a. *He ate his dinner after ___ got home.
   b. *When ___ arrived it was raining.

4.2.2 Stacking

Though there are restrictions (see Doherty 1993, 1994), some types of stacking of relatives involving SCR are acceptable. In (23), the SCR is followed by a regular subject relative (bracketed in the examples). See also (7h).

(23) a. I'm looking for someone speaks Irish well [who can do the translations].
   b. There's a woman lives in Ghent [who knows all about this stuff].
   c. I have a friend lives in Ghent [who knows all about this stuff].
   d. But there’s always something happens [that’s new or different]. (Guardian, G2, 18.4.2, page 4, col 3)

Prima facie it is hard to see how these would be derived in the topic-comment representation: the extraposed bracketed relative clauses in (23) should originate within S1, but would have to be extracted from S1 and end up somewhere to the right of the root domain S2. An analogous configuration involving a Hanging topic is ungrammatical, regardless of whether the relevant clause is restrictive (24a,b) or non-restrictive (24c,d):

---
9 There is speaker variation. For instance, Andrew Radford (p.c.) accepts this example.
(24) a. As for the course [which is taught in the second year], many students find it too hard.
   b. *As for the course, many students find it too hard, [which is taught in the second year].
   c. As for John’s syntax course, [which is taught in the second year], many students find it too hard,
   d. *As for John’s syntax course, many students find it too hard, [which is taught in the second year].

5 Conclusion

The background for this paper is the question to what extent syntactic structure can be taken to directly reflect interpretation, or, put differently, the degree to which information structure can drive syntactic representations. I have discussed and evaluated a topic-comment representation of SCR which closely reflects its discourse properties: a clause introducing a nominal constituent as its focus is topicalized and what looks like a relative clause is analysed as a non-embedded (and unembeddable) root clause providing a comment on the focussed nominal constituent. On closer inspection, though, such a representation has been shown to be problematic for both the internal and the external syntax of English SCR.

In a different context, the situation described here is arises with respect to the analysis of English *it* clefts, where the interpretive similarity with focus fronting in English has also led some researchers (Meinunger 1997, 1998, Frascarelli and Ramaglia 2013, and Sleeman 2011) to propose an analysis in which what looks like a relative clause, i.e. the cleft relative, is syntactically analysed as a matrix clause. Again, the root analysis of *it* clefts fails to predict specific properties of the internal and external syntax of clefts, a point made extensively in Haegeman, Meinunger and Vercauteren (2014).

References


1 The Object of Inquiry

Aspects of the Theory of Syntax (Aspects) is one of the defining texts of Generative Grammar (GG). Along with Logical Structure of Linguistic Theory (LSLT) (in its massively reduced Syntactic Structures [SS] avatar), Lectures on Government and Binding and the “Black Book,” The Minimalist Program (MP), Aspects defines the four major theoretical epochs of the Generative tradition. Looking back on it now, however, there is a sense in which Aspects constituted a misstep, albeit an extremely productive one. Its major theoretical innovation was ‘Deep Structure’ (DS). DS has three major properties: (i) it serves as the recursive engine of the grammar, (ii) it codes for (what we now call) ‘thematic information,’ and (iii) it is input to the transformational component that maps DS to ‘Surface Structure,’ the grammatical level that feeds the phonological component in the Standard (i.e. Aspects) theory. In the previous SS framework, the transformational component is the source of grammatical recursion. Aspects relocates this power to the base, context free phrase structure rules replacing embedding transformations as the engine generating bigger and bigger hierarchically structured phrase markers.

Variants of this Aspects picture persisted within GG syntactic theory until MP returned recursion to its earlier transformational home and argued for the elimination of Deep Structure (and all its Dish variants) to boot. In short, MP returned us to a version of the original LSLT/SS vision of how the grammar is organized. In this sense, the theoretical investigation that Aspects initiated, proved to be a wrong turn, given current assumptions.¹ Hence, theoretically speaking, syntax has rejected the Aspects conception of UG.

¹ Let me reiterate that this does not mean that it was an unproductive one. Theory in the thirty years between Aspects and MP was unbelievably fecund. Moreover, I believe that it could be argued (though I will not do so here) that by factoring grammatical operations between phrase structure rules and transformations allowed the latter’s properties to be brought more clearly into focus. Despite some very interesting attempts to re-unify structure building and movement operations in the MP period (via E/I-Merge), there remain important asymmetries between
Nonetheless, *Aspects*, in particular chapter 1, remains one of the most important documents in GG and is as relevant today as it was in 1965. Why? Because it is (one of) the best articulations to date of the subject matter of linguistics. Nothing better defines the goals of linguistic theory and outlines the kinds of theories we should be looking for. What *Aspects* did, and continues to do, is firmly link the aims of generative grammar with a Rationalist conception of mind and explanation (‘Rationalist’ here is in opposition to ‘Empiricist’). Its major achievement was to firmly ground the Generative enterprise in a Rationalist conception of mind and, more broadly, in a Rationalist conception of what constitutes a good scientific explanation. In what follows, I would like to highlight some of the ways that *Aspects* managed this.

2 Some Preliminaries

*Aspects* defines a primary object of study in linguistics to be the “underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance.” This “generative grammar” is a procedure for assigning interpretations for an infinite range of sentences (i.e. a procedure for relating a phonetic interpretation and a meaning over an infinite range of objects). So described, the first object of study is a generative procedure, a system of rules, a grammar that specifies sound/meaning pairs over an effectively unbounded domain.

Chomsky is similarly very explicit here that this conception is “mentalist” in that we are aiming to discern an “underlying mental reality” on which behavior supervenes. More pointedly, the object of study is not linguistic behavior itself (the “actual performance”) but a far more abstract substructure, which though used in the course of linguistic behavior is not identical to it. Thus, on this view, grammars are not summaries of linguistic behavior (i.e. generalizations over our comprehensions and productions or a model for a “speaker or a hearer” (9)), rather it is the abstract characterization of an abstract system of knowledge that gets put to use in various ways. Linguistic behavior is (at best) one possible source of evidence that we can use to probe the structure of the generative procedure, but grammars are in no sense summaries of behavioral patterns or representations of the regularities in the input.

This latter point is worth re-emphasizing in the current climate. In *Aspects* Chomsky proposes that we study not what people do linguistically but the underlying mental structure that describes a native speaker’s linguistic knowledge. To mark this point, *Aspects* distinguishes linguistic competence from performance and argues that a theory of the former is logically prior to a theory of the latter (viz. performance theories presuppose some account of competence). A theory of competence is a theory about a capacity, a theory of performance is a theory of how this capacity is exercised. In modern parlance, it roughly tracks the distinction between data structures and algorithms. Though there is an intimate relation between the two (and we can often learn a lot about each from the other), nonetheless, they are very different and confusion ensues if we don’t keep them apart.

*Aspects* introduces further distinctions to keep us from confusing these different though related domains. For example, it emphasizes (p. 11) the difference between ‘grammaticality’ and

---

2 As Katz and Bever (1976: 12) observe that “the most significant aspect of the transformational revolution is that it is a decisive defeat of Empiricism in an influential social science.”

3 Of course, how it gets used is an interesting empirical issue.
‘acceptability.’ The former is a property of phrase markers and their derivations. A given linguistic structure can be well formed (grammatical) or not (ungrammatical). So, the phrase marker of a given sentence might violate the Complex Noun Phrase Constraint. If it does, it is an ungrammatical structure (either cannot be generated or would be weeded out by some kind of filter). Native speakers may judge ungrammatical sentences to be unacceptable. And the fact that unacceptability is often a good indicator of ungrammaticality is what allows linguists to use acceptability judgments as empirical probes into grammatical structure. However, as Aspects notes, it is entirely possible that some unacceptable sentences are grammatical (e.g. Police police police buffalo buffalo buffalo) and that some ungrammatical sentences are acceptable (e.g. More people visited Rome than I did). Like the competence/performance distinction, the grammaticality/acceptability distinction is meant to guard us from mistaking the principle object of inquiry.

Aspects further identifies two different uses of the notion grammar. In the first instance a grammar describes the linguistic competence a native speaker of a given language has concerning that native language. Thus, the grammar of English would differ from that of French because native speakers of the two languages know different things (i.e. are in different mental states) reflecting the fact that they are competent speakers of different languages. However, there is another sense in which these native speakers have similar linguistic capacities for each could have acquired the other grammar but for adventitious circumstances. Thus, were Peter raised in Paris he would have acquired a French G and were Pierre raised in London he would have acquired an English G. Thus, despite their particular actual differences, Peter and Pierre’s capacity to learn either language (and of course, we should not stop with English and French, any natural language will do) is a common feature of our native speakers. And this higher order capacity is also a proposed object of linguistic study: what is the underlying capacity the native speakers have that allows them to acquire any G when exposed to uttered products of G? Aspects dubs this second capacity ‘Universal Grammar’ (UG) and contrasts it with ‘Grammar’ (G) tout court. It also outlines the conceptual dependencies between the two studies. A theory is dubbed ‘descriptively adequate’ if it faithfully describes a native speaker’s language particular G. A theory is explanatorily adequate if it shows how any particular G can be derived from the principles of UG given the kind of input native speaker’s are exposed to (the ‘Primary Linguistic Data’ [PLD]).

So to recap: Aspects identifies the object of study to be two related mental capacities. The first is the mental capacity that a particular G describes: the capacity to generate sound/meaning pairs for an unbounded number of linguistic structures. The second object of study is a second order capacity that UG describes: the capacity to derive first order grammatical capacities based on PLD. The first order capacity (partly) underlies our further capacity to engage in certain kinds of linguistic behavior (e.g. talking, comprehending, poetizing, explaining etc.). The second order capacity (partly) underlies the human capacity to become a native speaker of a natural language.

---

4 Though we tend to describe well formedness in dichotomous terms, there is nothing preventing us from taking sentences to have degrees of grammaticality, as Aspects (p. 11) notes.
3 The Rationalism of *Aspects*

There are several different Rationalist features encapsulated in the *Aspects* program. Let’s consider some of these in turn.

3.1 Capacities versus Regularities

Let’s first consider *Aspects*’ emphasis that what needs explaining is an abstract capacity rather than a form of behavior. Nancy Cartwright (1999) discusses the difference between capacities and regularities and makes the following point. The classical empiricist/Humean tradition rejected capacities/powers as occult residues of an earlier search for Aristotelian essences and insisted on founding all scientific knowledge on “the kinds of qualities that appear to us in experience (79)” (recall the dictum: nothing in the intellect that is not first in the senses!). Modern empiricists/Humeans endorse this antipathy to “powers” by treating the laws of nature as summaries of “what things do (82).” Cartwright contrasts this with the view that laws are about powers/capacities, which is not about what things do but “what it is in their nature to do (82).”

Here’s a quote that provides a good feel for what she has in mind (81-82):

> What we have done in modern science, as I see it, is to break the connection between what the explanatory nature is—what it is in and of itself—and what it does. An atom in an excited state, when agitated, emits photons and produces light. It is, I say, in the nature of an excited atom to produce light. Here the explanatory feature—an atom’s being in an excited state—is a structural feature of the atom… For modern science what something really is—how it is defined and identified—and what it is in its nature to do are separate things.

In short, there is an important metaphysical distinction that divides Empiricists and Rationalists. For the former the laws of nature are in effect summaries (perhaps statistical) of “actually exhibited behaviors”, for the latter they describe abstract “configurations of properties” or “structures.” These latter underlie, but are distinct from, behavior (“what appears on the surface”), these being “the result of the complex interaction of natures (81).”

Cartwright notes the close connection between the Rationalist conception of powers/capacities and the analytic method of inquiry characteristic of the physical sciences, often called “Galilean idealization.” She also provides several interesting reasons for insisting on the distinction between what something is versus what it does. Here are two.

First, given that visible behavior is an interaction effect of complex natures it is often impossible to actually see the contribution of the power/capacity one is interested in, even in the very contrived circumstances of controlled experiments. She illustrates this using Coulomb’s law and the interfering effects of gravity. As she points out:

> Coulomb’s law tells not what force charged particles experience but rather what it is in their nature, *qua* charged, to experience… What particles that are both massive and charged actually experience will depend on what tendency they have *qua* charged and what *qua* massive (82).
Thus, actual measurable forces are the result of the interaction of several powers/capacities and it takes great deal of idealization, experimentation, calculation and inference to (1) simply isolate the effects of just one and segregate it from everything else, viz. to find out how two charged bodies “would interact if their masses were zero.” And (2) to use the results of (1) to find out what the actual powers involved are:

The ultimate aim is to find out how the charged bodies interact not when their masses are zero, nor under any other specific set of circumstances, but how they interact qua charged.

Second, contrary to the accepted wisdom more often than not in the real world the same cause is not followed by the same effect. In fact, generating stable relations between cause and effect requires very careful contrivance in manufactured artificial experimental settings. Cartwright refers to these as nomological engines; set-ups that allow for invariant regular connections between what powers/natures/capacities can do and what they actually do. Except in such settings the Humean dictum that effects regularly follow causes is hardly apparent.

Outside the supervision of a laboratory or the closed casement of a factory-made module, what happens in one instance is rarely a guide to what will happen in others. Situations that lend themselves to generalizations are special… (86).

Cartwright’s discussion tracks the one we find in Aspects. The distinction between competence and performance is a rationalist one. Descriptions of individual grammars and theories of UG are intended to be accounts of human linguistic powers/capacities, not theories of linguistic behavior. Neither G nor UG is a summary of behavioral regularities (nor, for that matter, a summary of regularities found in the input). Indeed, as Aspects insists, linguistic behavior is a very complex interaction effect with competence being one of many (very poorly understood) factors behind it. The distinction between what a speaker knows (competence) and what a speaker puts this knowledge to use (performance) clearly echoes Cartwright’s rationalist themes. Similarly, the rejection of the idea that linguistic competence is just (a possibly fancy statistical) summary of behavior (or of the input) should be recognized as the linguistic version of the general Rationalist endorsement of the distinction between powers/capacities and their behavioral/phenomenal effects.

3.2 The Rich Structure of UG
Around the time that Aspects was written, linguistics was in the vanguard of the cognitive revolution, a major battle-ground on which Rationalism and Empiricism met and disputed. Aspects, with its focus on the necessity for a highly structured UG to account for how particular grammars are acquired, argued for a Rationalist understanding of minds against the then dominant Empiricist conceptions. Here’s what was taken to be at stake.

Empiricism, a species of environmentalism (natural selection being another), holds that minds are structured by the environments in which they are situated. Grammars, are, at best, “compressed memory representations of the regularities found in the input” (Lidz and Gagliardi

5 Cartwright observes that though doing (1) is difficult it is “just a stage; in itself this information is uninteresting.” (83-4).
2013). The leading metaphor is the mind as soft perfectly receptive wax tablet (or empty cupboard) which the external world shapes (or fills) via sensory input. The leading slogan, borrowed from the medievals, is “nothing in the intellect that is not first in the senses.” The mind, at its best, faithfully records the external world’s patterns through the windows of sensation. Good minds are good pattern matchers, able to track the generalizations in the data.6

Rationalists have a different animating picture. Leibniz, for example, opposed the wax tablet metaphor with another: ideas are in the mind in the way that a figure is implicit in the veins of a piece of marble (p. 52). The sculptor cuts along the marble’s grain to reveal the figures that are inchoately there. In this picture, the environment is the sculptor, the veined marble the mind. The image highlights two main differences with the empiricist picture. First, minds come to environments structured (“veined”). They have a natural grain, allowing some figures (ideas) to easily emerge while preventing or slowing the realization of others. Second, whereas a hot wax imprint of an object mirrors the contours of the imprinting object, there is no resemblance between the whacks of the chisel and the forms that such whackings bring to life. What’s in the senses may provoke the emergence of one or another mental structure, but not by summarizing the inputs into various generalizations, but in a more oblique way. Data is used to select among pre-specified options. Thus, linguistic input is to emerging mental structure as experimental data is to theory. It is used to test given hypotheses. Aspects’ elaborate (almost Bayesian) version of the acquisition problem (p. 31) presents a picture in which language acquisition requires the pre-specification of a set of (weighted/ordered) alternatives among which the data selects. Thus, Rationalists allow minds to represent external reality but deny that they do so in virtue of some sort of similarity obtaining between the sensory perceptions and the ideas they prompt. The mind is a selection device choosing among given alternatives, not a device for generalizing inputs into patterns. The metaphors are important here: whereas Rationalists postulated causal connections between mental content and environmental input they denied that environments shape those contents. The distinction between triggering and shaping is an important one.

Associationism is the modern avatar of empiricism. The technology is more sophisticated, neural nets and stimulus-response schedules replacing wax tablets and empty cupboards, but the guiding intuition is the same. Minds are pattern matchers able with sufficient exposure to the patterns around them to tune themselves to the patterns impinging on them. What made Aspects’ ideas about Generative Grammar so exciting was that they showed that this empiricist picture could not be right. To account for a native speaker’s linguistic competence requires that humans come equipped with highly structured special purpose mental procedures and this is inconsistent with empiricisms associationist psychology. Two features of linguistic competence were of particular importance: first that the competence emerges relatively rapidly, without the learning being guided and despite data that is far from perfect. Second, much of what speakers know about their language is not attested at all in the data they have access to and use. No data, no possible associationist route to the mind. Ergo: the mind must be structured.

3.3 Patterns and Generative Procedures
Aspects characterizes the acquisition problem as going from primary linguistic data (PLD) to grammars. PLD are “examples of linguistic performance” (25) while grammars “are systems of

---

6 See (i) above. Note that the relation between minds as pattern matchers lives comfortably with the rejection of the competence/performance distinction: regularities in behavior are what minds track and pattern detection is what minds do.
The Rationalism of Generative Grammar

rules” that constitute “the native speaker’s internally represented “theory of his language.”” (25).
The problem then is going from instances of used outputs of grammars to the rules that characterize (aka, generate) the linguistic objects used. Note the problem is not one of finding patterns in the data, but rules that generate it. Indeed, one of the interesting things about the project is that contrary to say Greenberg’s conception of Universals, the Aspects notion does not assume that there are patterns in the PLD to discern. The relation between data and theory is more remote than simple pattern detection. An analogy might help illustrate what this means.

Consider two kinds of patterns. The first kind is illustrated in sequences like (1):

1. (a) .222222…
   (b) .333333…
   (c) .454545…
   (d) .123412341234…

If asked to continue into the … range, a normal person (i.e. a college undergrad, the canonical psych subject and the only person buyable with a few “extra” credits, i.e. cheap) would continue (1a) with more 2s, (1c) with more 3s (1c) with 45s and (1d) with 1234s. Why, because the average person would detect the indicated pattern and generalize as indicated. People are good at detecting patterns of this sort. Hume discussed this kind of pattern recognition behavior, as have empiricists ever since. What the examples in (1) illustrate is constant conjunction, and this leads to a simple pattern that humans have little trouble extracting, (at least in the simple cases).

Now as we all know, this will not get us great results for examples like (2).

2. (a) .141592653589793…
   (b) .718281828459045…

The cognoscenti will have recognized (2a) as the decimal part of the decimal expansion of \( \pi \) (15 first digits) and (2b) as the decimal part of the decimal expansion of \( e \) (15 first digits). If our all purpose undergrad were asked to continue the series s/he would have a lot of trouble doing so (Don’t take my word for it. Try the next three digits). Why? Because these decimal expansions don’t display a regular pattern as they have none. That’s what makes these numbers irrational in contrast with the rational numbers in (1). However, and this is important, the fact that they don’t display a pattern does not mean that it is impossible to generate the decimal expansions in (2). It is possible and there are well known algorithms for doing so (as we display anon). However, though there are generative procedures for calculating the decimal expansions of \( \pi \) and \( e \), these procedures differ from the ones underlying (1) in that the products of the procedures don’t exhibit a perceptible surface pattern. The patterns, we might say, contrast in that the patterns in (1) carry the procedures for generating them in their patterning (Add 2,3, 45, 1234, to the end), while this is not so for the examples in (2). Put crudely, constant conjunction and association exercised on the patterning of 2s in (1a) lead to the rule ‘keep adding 2’ as the rule for generating (1a), while inspecting the patterning of digits in (2a) suggests nothing whatsoever about the rule that generates it (e.g. (3a)). And this, I believe, is an important conceptual fault line separating

---

7 This is based on a discussion in Berlinksy (1988).
8 There is surely a bound to this. Consider a decimal expansion whose period are sequences of 2,500 digits. This would likely be hard to spot and the wonders of “constant” conjunction would likely be much less apparent.
9 Answer: for \( \pi \): 2,3,8 and for \( e \): 2,3,5.
empiricists from rationalists. For empiricists, the paradigm case of a generative procedure is intimately related to the observable patternings generated while Rationalists have generally eschewed any “resemblance” between the generative procedure and the objects generated. Let me explain.

It’s uncontroversial that learners come to the task of language acquisition with biases. This just means that everyone agrees that what is acquired/learned is not a list, but a procedure that allows for unbounded extension of the given (finite) examples in determinate ways. Thus, everyone (viz. both Empiricists and Rationalists) agrees that the aim is to specify what biases a learner brings to the acquisition task. The difference lies in the nature of the biases each is willing to consider. Empiricists restrict the biases they are willing to entertain. The admissible biases are those that allow for the filtering of patterns from data. The leading Empiricist idea is that data reveal patterns and that learning amounts to finding these patterns in the patternings of the data. In other words, they picture the problem of learning as roughly illustrated by the examples in (1).

Rationalists are more catholic. Though they allow that Empiricist acquisition exists, they don’t restrict themselves to these. They allow that there are learning problems akin to that illustrated (2). And that this kind of learning demands departure from algorithms that look for “simple” patternings of data. In fact, it requires something like a pre-specification of the possible generative procedures. Here’s what I mean.

Consider learning the digital expansion of π. It’s possible to “learn” that some digital sequence is that of π by sampling the data (i.e. the digits) if, for example, one is biased to consider only a finite number of pre-specified procedures. Concretely, say I am given the generative procedures in (3a) and (3b) and am shown the digits in (2a). Could I discover how to continue the sequence so armed? Of course. I could quickly come to “know” that (2a) is the right generative procedure and so I could continue adding to the … as desired.

\[
\begin{align*}
(3a) \quad \pi &= 2 \sum_{k=0}^{\infty} \frac{k!}{(2k+1)!} = 2 \sum_{k=0}^{\infty} \frac{2^k k!^2}{(2k+1)!} = 2 \left[ 1 + \frac{1}{3} \left( 1 + \frac{2}{5} \left( 1 + \frac{3}{7} \left( 1 + \ldots \right) \right) \right) \right] \\
(3b) \quad e &= \lim_{n \to \infty} \left( 1 + \frac{1}{n} \right)^n = 1 + \frac{1}{1!} + \frac{1}{2!} + \frac{1}{3!} + \ldots
\end{align*}
\]

How would I come to know this? By plugging several values for K, n into (3a,b) and seeing what pops out. (3a) will spit out the sequence in (2a) and (3b) that of (2b). These generative procedures will diverge very quickly. Indeed the first computed digit renders us confident that asked to choose (3a) or (3b) given the data in (2a), (3a) is an easy choice. The moral: even if there are no patterns in the patternings of the data acquisition via data sampling is possible if the range of relevant choices is sufficiently articulated and bounded.

This is just a thought experiment, but I think that it highlights several features of importance. First, that everyone is knee deep in given biases, aka: innate, given modes of generalizations. The question is not whether these exist but what they are. Empiricists, from the Rationalist point

---

10 Hence the ton of work done on categorization, categorization of prior categorizations, categorization of prior categorizations of prior categorizations…

11 Or may exist. Whether it does is likely more complicated than usually assumed as Randy Gallistel’s work has shown. If Randy is right, then even the parade cases for associationism are considerably less empiricist than often assumed.
of view, unduly restrict the admissible biases to those constructed to find patterns in the patternings of the data. Second, that even in the absence of patterned data, learning is possible if we consider it as a choice among given hypotheses. Structured hypothesis spaces allow one to find generative procedures whose products display no obvious surface patterns.

A historical aside: Here, Cartwright provides the ingredients for a nice reconstructed history. Putting more than a few words in her mouth, it would go something like this:

In the beginning there was Aristotle. For him minds could form concepts/identify substances from observation of the elements that instanced them (you learn ‘tiger’ by inspecting tigers, tiger-patterns lead to ‘tiger’ concepts/extracted tiger-substances). The 17th century dumped Aristotle’s epistemology and metaphysics. One strain rejected the substances and substituted the patterns visible to the naked eye (there is no concept/substance ‘tiger’ just some perceptible tiger patternings). This grew up to become Empiricism. The second, retained the idea of concepts/substances but gave up the idea that these were necessarily manifest in visible surface properties of experience (so ‘tiger’ may be triggered by tigers but the concept contains a whole lot more than what was provided in experience, even what was provided in the patternings). This view grew up to be Rationalism. Empiricists rejected the idea that conceptual contents contain more than meets the eye. Rationalists gave up the idea the content of concepts are exhausted by what meets the eye.

The Empiricist/Rationalist distinction noted here is reflected in different ways of understanding ‘universals.’ In Aspects, ‘universal’ means a feature of UG, the function that takes you from PLD to a particular G. There is every reason to think that these universals are very abstract and describe the class of generative procedures that humans can attain. There is no reason to think that a principle of UG will generate any particular patterns within the phrasal products of a particular G. ‘Universals’ of UG are not shared properties displayed in the products of every G. They are restrictions on the class of admissible operations and, if “visible” at all, will only leave a mark on what is excluded from the language. Surveying the well formed products of any particular G will tell you very little about the generative procedure that gives you these products.

A useful contrast to the Aspects notion of ‘universal’ is Greenberg’s. Here a linguistic pattern is universal if it is visible in the surface patterns of every relevant language, e.g. all languages contain nouns and verbs or, if a language is SVO then it is prepositional. Greenberg universals are claims about the patterning of linguistic observables and will be apparent from well-formed instances of the language (e.g. the acceptable strings). Aspects universals are claims about generative procedures, not products. These are two very different things.

We can go further: not only need universals not be “visible” in the products of Gs they need not be visible across all Gs. UG in the Aspects sense does not require that every G have common rules or categories or operations. UG delimits the class of possible generative procedures not the specific rules they contain nor the linguistic products generated. So, ‘universals’ in Aspects don’t

---

12 The “debate” over the relevance of Piraha for the theory of universal grammar revolves around confusing these two sense of ‘universal.’ Everett takes it to be an implication of the claim that grammars are recursive that every language display unbounded embedding. He claims that Piraha limits embedding to degree 1 and concludes that recursion is not a property of UG. Everett’s point makes sense if one understands ‘universal’ in the Greenberg sense of the term, not the sense of the term in Aspects.
denote the class of admissible visible linguistic patterns, and so we do not expect to be able to find these universals by inductive examination of the (used) outputs of these procedures any more than we would expect to be able to induce (3a) by examining integer patterns in (2a).

3.4 The Possible and the Actual
Rationalists explain the actual in terms of the possible. This is true both ontologically and epistemologically. The actual, what occurs, is the combination of contingent initial conditions and non-contingent “laws.” Furthermore, the operative laws causally at play in any particular case display complex interaction effects depending on the specific initial conditions at hand. Thus, what one sees or measures is the result of at least two kinds of contingencies. The aim of experiment and theoretical inquiry is to disentangle the principled laws from the contingent initial conditions and their resultant interaction effects. As Cartwright put it discussing the interaction of charged particles:

The ultimate aim is to find out how the charged bodies interact not … under any other specific set of circumstances, but how they interact qua charged.

The aim of explanation focuses on what’s possible (“qua charged”) not what one sees in any given experiment, the latter being the product of “specific” circumstances and thus obscuring the workings of the general laws. As the aim is to understand these laws, a good experiment allows one to see through the specific details to the mechanisms that define the class of possible effects, the particular one at hand being merely contingent.

We can put this another way: what’s real is the class of possible effects. What we actually see is derivative on these laws and some historical accidents.

This fits very well with the conception in Aspects. Competence is principled. Performance is not. Gs are theories of competence. They characterize a set of expressions far greater than the particular set of linguistic objects that any (indeed, all) native speakers can ever possibly actually encounter. Further, any given performance (e.g. produced or perceived utterance) is a massive interaction effect of various mental modules, only one of which is the grammar. Plus there are entirely adventitious initial conditions. What is uttered is contingent on the specific circumstances in which the utterance is produced. They are caused by a variety of extremely poorly understood factors (e.g. the “free” decision to utter the utterance). In sum, Gs, which characterize the class of possible sentences of a given language, are considerably less contingent on initial conditions or the effects of other mental modules than the utterances actually produced. It is in this sense that Gs are more real than the utterances they (partially) characterize.

The G any given native speaker has is also contingent, this time on the linguistic environment the speaker grew up in. That I happen to speak “English” is an accident. That I have acquired a G with the characteristic properties that UG demands is not.

In sum, the rationalist perspective in Aspects enjoins us to abstract away from these contingencies in describing particular Gs (the theory of the possible linguistic objects in a given language) and UG (the theory of the possible Gs available to humans as such). UG and the Gs that it permits are ontologically more real (i.e. less contingent on accidents) than the actual Gs that arise and the products that one actually encounters and produces.

Aspects also makes these Gs and UG epistemologically more basic as well. This is particularly clear in the case of UG. The Aspects project assumes that the class of acquirable Gs
is severely circumscribed. The hypothesis space is circumscribed and the weighting of alternative grammars is highly specified. The role that PLD plays in the acquisition process, therefore, is pictured as being relatively trivial: it selects among a relatively small set of alternatives. Understanding language acquisition is largely understanding how the space of possibilities is configured and how the options are weighted. Thus, epistemologically, UG is more basic than the Gs speakers happen to have acquired in being a pre-condition for their acquisition.

Similarly for Gs. As noted they define an (effectively) infinite class of linguistic structures only a finite number of which are actually performed. What you “know” is a precondition of what you “do.” The possible circumscribes the actual.

4 Conclusion

Aspects is not only a seminal document within modern linguistics, it is one of the intellectual struts of the cognitive revolution. Chapter 1 makes clear the connection. So, despite the fact that the central theoretical features advanced in Aspects have since been rejected (or considerably modified), chapter 1 remains a founding document of the Generative enterprise. Why? Because it is the first (and still best) concise and articulate outline of the Rationalist program in linguistics. We have learned a lot about Gs and UG since Aspects. However, the general program outlined there remains as exciting as ever.

References

ASPECTS AND BEYOND: THE CASE FOR GENERALIZED CONSTRUCTION

Jan Koster
University of Groningen

1 Aspects and the Panglossian Fallacy

When Aspects appeared in 1965, it was naturally seen against the backdrop of issues raised by Syntactic Structures (Chomsky 1957). For those who entered the field in the late 1960s, like myself, the two classics were studied more or less in tandem. Syntactic Structures was widely seen as the founding document of the revolution and Aspects was optimistically received as the next great leap forward. Aspects added an exciting philosophical-historical dimension by situating generative grammar in the tradition of rationalistic epistemology. At the same time, empiricists like Skinner and Quine were attacked, giving the whole discussion a dimension of contemporary urgency. This was true even for Europe, where often phenomenologists were targeted instead of behaviorists (as in Staal 1967) because it was hard to find a living behaviorist. The philosophical critique also had a biological dimension, with approving references to ethology (Chomsky 1959). This field, created by Konrad Lorenz and Niko Tinbergen, was later seen as a precursor of sociobiology.

At a more technical level, the appearance of Aspects was hailed as meeting the widely felt need for more semantics. A lexical component was added to generative grammars, with a discussion of selection restrictions that explored the scope of features as made popular by Katz and Fodor (1963). The “strings underlying kernel sentences” of Syntactic Structures, together with some other adjustments, were re-baptized as “deep structures.” This suggestive term gave the field an enormous public appeal at the time. This is especially true as the notion was further explored along the lines of Katz and Postal (1964), who had claimed that Deep Structure was the level at which all grammatical meaning was represented. The Katz/Postal hypothesis gave soon rise to Generative Semantics and related “linguistic wars” (Harris 1993), which is a topic that I will leave gladly aside.

Instead, I would like to focus on the consequences of Aspects’ all-import addition of a lexicon to generative grammars. It is my claim that there is a serious tension—unresolved up until the present day—between the psycho-biological perspective introduced in chapter 1 of
Aspects and the lexicalism that followed from chapter 2. The basic conflict is very simple: lexical items, with whatever properties they have, are ultimately invented socio-cultural objects, rather than something individual-psychological or biological. As invented cultural objects, lexical items belong to what Saussure called a “trésor commun.” For lexical items (and their properties) to qualify as non-individual cultural objects, it is irrelevant that they depend on our individual psycho-biological capacities, for the elementary reason that all cultural objects are dependent on these capacities. A musical composition, for instance, can only function as such thanks to capacities (or a combination of capacities) that are unique to our species. In other words, Aspects introduced a deep confusion into our field by failing to make the obvious distinction between biology and applied biology. Failing to make this distinction was named “panglossianism” by Gould and Lewontin (1979), in a context of similar discussions about sociobiology.

The panglossian fallacy in combinatorial syntax can only be avoided by abstracting away from cultural objects and their properties as found in the lexicon. This is precisely what is done in Syntactic Structures (pre-Aspects) and in most current forms of Minimalism (post-Aspects). In Syntactic Structures, a lexicon is simply missing. In minimalist grammars, the operation Merge generates structures independently of lexical properties as part of a “Galilean-style” interaction between abstract perfections (like Merge) and the “disturbing” imperfections as found in the lexicon (and more generally, at the “interfaces”).

2 The Redundancy Problem and Generalized Construal

Another persistent problem caused by Aspects’ introduction of a lexicon is known as “the redundancy problem.” Compare an Aspects-style PS-rule (1) and an Aspects-style lexical entry (2):

(1) VP → V NP

(2) catch : [+V, --NP]

It appears that the lexical specification of a transitive verb like catch mimics the PS-rule (1), making the latter superfluous. It is more economical to “project” a structure like [VP V NP] directly from the lexical entry (2), as it eliminates the need for (1). This was, in fact, one of the rationales behind the development of X-bar theory since Chomsky (1970), which was explicitly motivated by the redundancy problem (see Chomsky 1981: 31ff).

Actually, (2) is a very minimal specification of the structure entailed by a verb. In an extended sense, the verb is the lexical head of an entire clause, with the functional domains as completely predictable additions. In accordance with current theories, a verb like catch entails the following structure:

(3) catch : [+V, [CP (DP) [C [TP DP [T [VP V [VP -- DP]]]]]]]

Since a sentence structure is entirely predictable from the verb, the only economical way to generate a sentence is by spelling out the structural potential of the verb (as in (3)) together with the selection of the verb itself. The same can be said about other lexical categories and we must conclude at this point that syntactic structures are the spelled out properties (“templates”) of the cultural objects known as lexical items.
As in the case of the panglossian fallacy, we should ask ourselves now to what extent the redundancy problem is still with us in minimalist syntax. One goal of minimalism is to isolate the more abstract, less stipulative fundamentals hidden in the web of stipulations coming with lexical templates like (3). So far, the operation Merge is seen as the Galilean Grail of this enterprise. Binary Merge takes two linguistic objects, X and Y, and combines them to the set \{X, Y\}. As lexical items not only include morphemes and words but also the results of earlier applications of Merge, the operation is recursive. In short, Merge creates binary, hierarchical structures with recursion but without the stipulated linear order we assumed for lexical templates like (3).

Obviously, the redundancy remains if Merge is seen as a structure-generating operation. Like the PS-rules in *Aspects*, Merge introduces the hierarchical structure with recursion that is also represented by lexical templates like (3). In a system with Merge, the redundancy can only be avoided by denying properties like in (3) to lexical items, an assumption that would be at variance with the empirical facts. I therefore conclude that Merge (as a structure-generating operation) must be rejected, particularly on the basis of the redundancy problem. As in pre-minimalist days, the only known non-redundant procedure of sentence generation is by spelling out lexical templates directly. Thus, when a verb like *catch* is selected from the lexicon, the template structure in (3) comes along automatically. Further lexicalization of the template may conclude the sentence generation process.

In an important sense, my negative conclusion about Merge is unfortunate, as Merge seems to implement some of what I see as the right properties. Apart from hierarchical structure with recursion, Merge also involves binarity and locality. Consider the most commonly accepted output of Merge, with X and Y as the terms merged and Z a label:

\[(4) \{Z \{X, Y\}\}\]

The structure is binary in that it involves exactly two merged terms, X and Y. It is strictly local in the sense that it does not involve variables: X and Y are “sisters” exhausting their hierarchical level. The label Z is one level up and exhausting the next level together with the complex object \{X, Y\}. If binarity and strict locality are desirable properties, how can we preserve these “good” properties of Merge without falling back into the redundancy trap of *Aspects*?

Luckily, there is a very simple solution to this problem. The problematic redundancy can be avoided by generating sentences not by Merge but by lexical templates like (3). The relevant properties of Merge can be preserved by giving up its status as a sentence-generating vehicle and reinterpret (4) as a meta-theoretical constraint on possible syntactic structures (including lexical templates). This constraint-based approach neatly accounts for the hierarchical structure, the possible recursion and the binary character of structures like (3). More about locality in what follows.

In contrast with Merge, the constraint-based approach not only solves the redundancy problem, it also happens to have unexpected unifying potential with respect to the rest of syntax. It is my claim that (4), adjusted and reformulated as a constraint, unifies the properties of base rules (i.e., template activations), traditional movement rules and rules of construal (agreement, anaphoric binding, etc.).

Consider a typical construal like anaphor binding:

\[(5)\text{John saw himself.}\]
This construal happens to be binary in the sense of (4), with John as the X-term and himself as the Y-term. A prima facie complication is that X and Y are “sisters” in (4) but not in (5), as (5) has the following structure:

\[(6) \left[ Z \ X \ [\text{saw} \ Y] \right] \]

X is not adjacent to Y but separated from it by saw, and more generally, by a structural segment of variable size. The VP [saw himself], however, is in full agreement with (4), with saw as X-term and himself as the Y-term. In general, construals have the following format (X the antecedent of Y in some local domain Z):

\[(7) \left[ Z \ldots \ X \ldots \ Y \ldots \right] \]

The output of traditional (binary) PS-rules (or the corresponding parts of X-bar schemata) is as follows (with (4) as its minimalist set-based translation):

\[(8) \left[ Z \ X \ Y \right] \]

The formats (7) and (8) are close enough to try a full-fledged unification. The main difference is that (7), unlike (8), contains variables (as indicated by the dots (…)). Elimination of the variables in (7) appears to be surprisingly simple, as I will show shortly. I will claim that not Merge but the triadic structure (8) is the Galilean Grail we are looking for in language. Interpreted as a meta-constraint on possible structures, (8) preserves some of the crucial properties of Merge but has a scope far beyond it, as suggested above. It not only constrains base structures, but also the traditional construals and movement rules. I will refer to this theory as Generalized Construal.

### 3 Triads and Their Functions

In biological and cultural structures a common distinction is made between form and function. Formally speaking, the triad (8) defines the configurational matrix of strict locality, which I consider the abstract formal core of syntax. Strict locality (in (8)) means that X and Y can only be directly related to Z or to each other. Relatedness beyond the confinement of the triad is only possible by mediation of Z, which can be the sister of some X or Y in the next triad up:

\[(9) \left[ Z' \ X' \ [Z \ X \ Y] \right] \]

Thus, X and Y cannot be directly related to X’ but only through the “sister” Z of X’. A syntactic structure in natural language consists of one or more triads, where the latter define the strictly local paths (of one or more links) required for syntactic interaction. Following Jan-Wouter Zwart (2011) and basing myself on the pervasive asymmetry of syntactic relations, I tend to assume that X and Y do not form a set (as in Merge) but an ordered pair. Formally, then, triadic constraints like (8) are labeled ordered pairs. It should be emphasized that labels (Z) are absolutely essential in this conception, as they mediate between successive triads. Labels, once more, are the “escape hatches” that make strict locality compatible with relations over some distance.

So far the formal aspects of triads. As for how they function, we have to look at the
content of lexical elements. It appears that all local syntactic activity involves construal of the same type, namely “action” driven by incompleteness. Consider again a construal as in (5), repeated here as (10):

(10) John saw himself.

The anaphor himself is incomplete in that it misses what is traditionally called a referential index i. This problem can be remedied by sharing the referential index of John. This kind of property sharing in a local domain, I consider the essence of construal (see Koster 1987: 8ff), which functions within the formal limits set by the triadic configurational matrix (8). In terms of (8), John is X and himself is Y. Since the X and Y in question are not sisters, we have a violation of (8) at first sight. Here, however, comes percolation to the rescue. Percolation is the traditional umbrella term for “upward” projection phenomena, which might subsume the labeling algorithms discussed in the current minimalist literature. These algorithms are mainly concerned with headedness and endocentricity issues, but it is clear that there is a wider range of cases in which structures from below determine the character of structures above them. An example that comes to mind is the formation of complex Wh-phases, as in Pied Piping: [which], [which man], [of which man], [the father of which man], [with the father of which man], etc. In such examples, the Wh-feature is inherited by ever more inclusive structures.

The key idea leading to unification in terms of strict locality is the assumption that “incompleteness” is an inheritable feature under percolation. Thus, we do not need special stipulations to establish that a VP that contains an incomplete category is itself also incomplete in some sense. Let us assume that /F is a notation indicating that a category is incomplete with respect to feature(s) F (notation inspired by Gazdar 1981). Then, without instantaneous satisfaction, the following holds:

(11) [F ... /F ... ]

Applied to (10), that leads to the following representation:

(12) [John, [VP/i saw himself/i ]]

This representation says that himself/i is incomplete with respect to referential index i. This incompleteness is inherited by the VP, written as VP/i. The incomplete VP/i can be completed by sharing the index i of its sister John. In other words, completion by property sharing can be “postponed” by the presence of a path made up from strictly local steps, in accordance with (8) (with the path theory not only inspired by Gazdar but also by Kayne 1984). As demonstrated by this example, strict locality means that variables are eliminated from construals.

4 External and Internal Merge as Construals

In standard minimalist theories, base structures and displacement (“movement”) constructions are formed by ad hoc stipulations, formerly known as Merge and Move and currently referred to as “External Merge” (EM) and “Internal Merge” (IM). Both are ad hoc, but IM is even less convincing than EM, as it leads to the utterly problematic “copying theory of movement.” A
copying theory would be exceptional in that it would involve the interaction of equally (in)complete categories, while, as we have seen, syntactic interaction in general is driven by the relative incompleteness of one of the two interacting categories. At least as bad is the further complication that copies must be distinguished from unrelated, accidental repetitions of the same category. One can wishfully think that the damage of further stipulations is limited by the fact that the interacting categories are in the same “phase”, but that does not seem to bring the desired Galilean perfection as close as one might hope for.

Luckily, we do not have to answer question as to the relative perfection of EM and IM, because there is no reason to assume that Merge exists in the first place. Both EM and IM refer to facts that are to be seen as entirely straightforward instances of Generalized Construal as explored in this article. Consider a Verb-Object construction (I use Dutch OV order for reasons of exposition):

\[ (13) \text{[VP DP V]} \]

You can see this structure as the result of a derivation involving Merge of DP and V, with label V:

\[ (14) \text{Merge (DP, V) = \{V \{DP, V\}\}} \]

However, such a rule is entirely superfluous. Not only are syntactic structures generated by spelling out lexical templates, moreover these templates owe their properties to the general constraint on construals (8), which has a much broader scope than Merge. A structure like (13) just shows a regular construal, with DP as the X-term and V as the Y-term. A transitive verb is incomplete with respect to an object DP, represented as V/DP:

\[ (15) \text{[VP DP V/DP]} \]

The incomplete V/DP is completed by its sister DP. This procedure not only makes phrase structure a regular instance of construal, it also makes the non-trivial prediction that completion by complements can be postponed in principle. The ubiquitous scrambling in many SOV languages fits the bill. Thus in Dutch, an object can be separated from its verb by an arbitrary number of adverbials, including adverbials that are traditionally seen as being higher than the VP:

\[ (16) \text{dat Jan [het boek [waarschijnlijk [ -- [gisteren [VP -- las]]]]]} \]

\[ \text{that John the book probably yesterday read} \]

\[ \text{“that John probably read the book, yesterday”} \]

The object (\textit{het boek}) can also occupy both positions indicated by --. Since the verb is transitive, it is incomplete (V/DP). Completion can be immediate, as in (15), or it can be postponed, as the result of the percolation of the incompleteness feature /DP:

\[ (17) \text{[DP [DP waarschijnlijk [DP gisteren [V/DP las]]]]} \]

Completion of /DP is done in exactly the same way in (15) as in (17), namely by a DP sister and in accordance with (8). Note that scrambling in Dutch does not create A-bar positions (as one would expect under “movement”). It has been recognized for a long time that scrambled DPs
have the properties of A-positions in Dutch, confirming their status as base-positions (see Vanden Wyngaerd 1989 for discussion).

One of the biggest obstacles for unification along the lines of Generalized Construal has been Chomsky’s early rejection of it (see Chomsky 1973: 284). The attempts to let dislocation (“movement”) stand out as something special has led to influential but superfluous ad hoc extensions of grammar, as manifested by concepts like “Move Alpha” and “Internal Merge.” It is easy to see, however, that movement constructions are regular instances of Generalized Construal. Consider the following sentence:

\[(18) \text{What did Mary catch DP?}\]

This sentence can be generated as follows. Selection of the verb *catch* activates its lexical template (3):

\[(19) \left[CP (DP) [C [TP DP [T [vP v [vP catch DP ]]]]]\right]\]

Further lexicalization involves some non-trivial problems, such as the general requirement that there must be a one-one match between functional elements (such as argument DPs) and their corresponding lexical contents. Ignoring these problems here, (19) can be further lexicalized, for instance resulting into the structure underlying (18):

\[(20) \left[CP \text{what} [ \text{did} [TP \text{Mary} [T [vP v [vP catch DP ]]]]]\right]\]

This structure is incomplete at first sight in that the rightmost DP is not directly lexicalized. As in the other cases we have discussed, incompleteness is upwardly inherited (percolated from label to label), eventually leading to completion by the leftmost lexical content, i.e., the DP *what*. The percolation structure is as follows (with /DP standing for incompleteness with respect to some lexical content DP):

\[(21) \left[CP \left[DP \text{what} \right] /DP \text{did} [TP/DP \text{Mary} [DP/DP T [vP/DP v [vP/DP catch DP ]]]]]\right]\]

Like all “movements”, this is an entirely regular case of postponed completion. The lexical content of the DP *what* is shared by the object of *catch* and the two positions are connected by a completely regular path permitted by a sequence of triads with the form specified by (8).

In short, the Generalized Construal approach makes “movement” an ordinary, regular case of postponed completion, while the approach based on Internal Merge and copying let it stand out as something anomalous. With IM and copying, displacement would be the only grammatical dependency based on identity (of the two copies) rather than on the non-identity that usually determines the need for completion.

### 5 Formal and Lexical Locality

In conclusion, I would like to make a few remarks about the difference between triadic (“strict”) locality and other forms of locality, like those found in Ross’s island conditions (Ross 1967), c-command (Reinhart 1976), the classical binding theory (Chomsky 1981) and phase theory
It has received little or no attention that c-command is crucially different from the other locality principles. C-command is formulated in purely formal terms, while the other conditions refer to specific, lexically-based categories, like CP, DP, vP, etc. From a “Galilean” perspective, then, c-command points in the direction of a structural level deeper than that covered by the other locality principles, as the latter are not purely formal but the result of application to lexical material. Applications always involve human culture, while the purely formal is pre-application and therefore providing a window on an older and deeper layer of structure.

The configuration matrix (8), then, is a generalization of the purely formal c-command pattern and not the result of lexical application, like the other locality principles. To clarify this, let us repeat (8) here as (22):

\[(22) [z X Y ]\]

C-command is usually seen as an asymmetrical relation. The configurational matrix (22), however, is completely symmetrical with respect to c-command: X c-commands Y and Y c-commands X. However, at the less abstract, applied level there is a functional asymmetry between X and Y. One of them, for instance Y, can be incomplete (dependent) while X is the completing (independent) term:

\[(23) [z X Y/X ]\]

This is the asymmetry that was briefly discussed above, with reference to Zwart (2011). In actual languages, this leads to a left-right asymmetry: in the majority of “real life” cases, the completing term X is to the left of the dependent, incomplete term Y/X. From a minimalist perspective, this symmetry-breaking result is an imperfection, probably due to the performance factors that determine the left-right organization of speech. Consider, for instance, “filler-gap” constructions. In principle, the filler could be on either side of the gap, but in practice the filler is usually to the left of the gap. A gap-first configuration would probably be more costly from a memory point of view, as the appearance of the gap would involve a postponed resolution by the filler. A filler-first configuration, however, allows instantaneous resolution as soon as the gap appears. These performance factors presumably lead to the apparent asymmetry with respect to c-command. At the deepest level, however, locality (as expressed by (22)), can be kept completely symmetrical.

For reasons of space, I will not say much here about the other locality principles (islands, binding domains, etc.). From the present perspective, the main issue is as follows. Although triads limit possible syntactic interactions to the strictly local configurations specified by (22), “escape” via Z and the construction of percolation paths makes it in principle possible to have paths of unlimited length. This, then, must be true for the deepest, purely formal level. At the application level, however, where (22) is implemented via lexical categories, paths appear to be of limited length. For most construals (binding, agreement, etc.) paths are limited to some clause type. For displacements, with some marked exceptions, the islands appear to be extended maximal phrases (see Koster 1978, 1987). Once more, this lexical locality seems to be an imperfection due to memory organization, as purely formal locality (22) allows unlimited paths by the iteration of strictly local percolation steps.

Whatever will turn out to be the correct theory of lexical locality, the existence of a
deeper level of strict, purely formal locality has an interesting consequence: it radically eliminates variables in the sense of Ross’s title Constraints on Variables in Syntax (1967). Each stretch of structure separating two terms X and Y must be reducible to a chain of one or more, variable-free triads. This allows us to eliminate the dots from (7) and completes the full unification of all construals in accordance with the format of (8) (= (22)).

6 Conclusion

Of the rich legacy of Aspects, I have highlighted its most important innovation, namely the addition of a lexicon to generative grammars. The introduction of a lexicon led to two issues that are unresolved until the present day. The first problem is that lexical items, no matter their biological basis, are humanly invented cultural objects that, as such, belong to what Saussure called a “trésor commun.” It is therefore an error to see language, even in the narrowest sense, in purely biological or individual-psychological terms. In abstraction from the invented lexicon, the underlying faculties, no matter how innate, have no proven linguistic function. Language is not a set of biological structures but the application, among other things, of a set of biological structures. Application means function assignment by human agency, the functionality to be preserved in a common, public culture. So far, talk about an individual “faculty of language” has not appeared to be more than a misleading panglossian way of speaking.

Another consequence of Aspects’ introduction of a lexicon was the growing insight that syntactic structures are properties of lexical items, best projected directly from the lexicon rather than by redundant phrase structure rules. This was well-understood in Chomsky (1981) but somehow the insight got lost in minimalism. Up until the present day, it is unclear how Merge can be formulated without introducing properties that already “exist” as properties of the elements to be merged. I therefore propose to maintain the GB idea that sentence generation is the spelling out of lexical properties. Merge does not exist under this proposal.

In order to maintain the “good” properties of Merge, I propose to reformulate it as a meta-constraint on possible syntactic structures (including lexical templates). Functionally, this establishes a configurational matrix for syntactic interactions under the operations “share” and “percolate.” The resulting triadic structure of syntactic interaction entirely eliminates Ross’s variables from the rules of syntax. Most important of all, the proposed theory of Generalized Construal unifies base rules (Merge), movements and various other construals under a common set of properties.

References


ASPECTS OF THE THEORY OF PHRASE STRUCTURE

HOWARD LASNIK
University of Maryland

1 Generalized Transformations vs. Recursion in the Base

There is a certain irony implicit in the Chomsky (1957) argument for the necessity of $\Sigma$, $F$ (context free rewriting) grammars. Chomsky shows straightforwardly that formal languages with unbounded mirror image properties go beyond the bounds of finite state Markov description. Then, upon suggesting that human languages have just such properties, he proposes $\Sigma$, $F$ grammars as the necessary extension, since such grammars are admirably well suited to characterizing mirror image languages. The irony resides in the fact that the theory of human language grammar that Chomsky then presents, following the one articulated in great detail in Chomsky (1955), restricts the power of the $\Sigma$, $F$ module in precisely such a way that it cannot handle the phenomena at issue. In particular, the theory explicitly and completely disallows recursion in this module (pp. 517-519). In this model, the infinitude of human languages is the responsibility of generalized transformations - operations melding separate phrase markers together into one phrase marker.

Chomsky (1965) resolves this irony by removing the restriction on the $\Sigma$, $F$ module, and thus allowing recursion ‘in the base.’ Though Chomsky (1955) had indicated that the restriction constituted a simplification, he didn’t actually offer any arguments to that effect. One might actually argue that removing this restriction is a simplification. After all, it seems to be a stipulation. Further, it is not trivial to determine whether there is recursion in the base. Certainly, the determination is trivial if there is a rule like (1):

(1) \[ A \rightarrow BA \]

But recursion might involve a pair of rules rather than any one rule:

(2) \[ A \rightarrow BC \]
    \[ C \rightarrow DA \]
Or a trio:

(3)  
   A → BC  
   C → DE  
   E → FA

In fact, there is no limit on how large the minimal group of rules might be that yield recursion.

Chomsky (1965) offered two major arguments for the revised approach (a major part of what he dubbed the ‘standard theory’). First, Chomsky claimed that the theory of transformational grammar is simplified by this change, the simplification being that the notion “generalized transformation” is eliminated entirely, at no apparent cost. Thus, in place of three syntactic operations - Σ, F rules, singulary transformations, and generalized transformations - we have just the first two. Further, the construct “Transformation-marker” is eliminated, as its major work was to show exactly how the separate trees combine into one, but now that is transparently represented in the initial phrase marker, the ‘deep structure.’

Chomsky's second argument is even more interesting. He argues that while there is extensive ordering among singulary transformations (situations where a derivation produces an unacceptable sentence if two transformations are applied in reverse order), “there are no known cases of ordering among generalized embedding transformations although such ordering is permitted by the theory of Transformation-markers” (1965: 133). Further, while there are many cases of singulary transformations that must apply to a constituent sentence before it is embedded, or that must apply to a ‘matrix’ sentence after another sentence is embedded in it, “there are no really convincing cases of singulary transformations that must apply to a matrix sentence before a sentence transform is embedded in it…”

In both frameworks, the set of singulary transformations was seen as a linear sequence: an ordered list. Given the Aspects modification, this list of rules is claimed to apply ‘cyclically,’ first operating on the most deeply embedded clause, then the next most deeply embedded, and so on, working ‘up the tree’ until they apply on the highest clause, the entire generalized phrase marker. Thus, singulary transformations apply to constituent sentences ‘before’ they are embedded, and to matrix sentences ‘after’ embedding has taken place. “The ordering possibilities that are permitted by the theory of Transformation-markers but apparently never put to use are now excluded in principle” (1965: 135).

It is important to note that within the Minimalist program, Chomsky argues against a level of deep structure, and Chomsky (1993) argues for generalized transformations as the sole structure-creating operation, responsible even for the structure of single clause sentences. Furthermore, singulary transformational operations are interspersed with these generalized transformational operations, again roughly as in the much earlier model. But what of the powerful Chomsky (1965) argument against such a model, that it allowed derivations that never actually occur in human languages. The model with recursion in the base excluded those unwanted derivations. However, on closer inspection, it was not actually elimination of generalized transformations that had this limiting effect. Rather, it was the stipulation that transformations operate strictly cyclically, starting on the most deeply embedded clause and proceeding monotonically up the tree. Chomsky (1993) observes that a condition with the same effect can be imposed on the operation of generalized transformations and their interaction with singulary transformations. This condition, often called the ‘extension condition,’ simply requires that a transformational operation ‘extends’ the tree upwards. This guarantees the same sort of
monotonic derivations as those permitted by Chomsky (1965).

The one remaining Aspects argument against generalized transformations can also be straightforwardly addressed. Chomsky had argued that eliminating generalized transformations yields a simplified theory, with one class of complex operations jettisoned in favor of an expanded role for a component that was independently necessary, the phrase structure rule component. This was a very good argument. But now that 1965 argument can be reversed on itself: Eliminate phrase structure rules!

2 Redundancy Between Subcategorization and Phrase Structure (PS) Rules

Aspects also contains the germ of another argument for the elimination of phrase structure rules. Chomsky (p. 96) discusses the phrase structure rules for VP, such as the following, with a representative VP using each rule:

(4)  
(i) VP \rightarrow V  
  (elapse)
(ii) VP \rightarrow V NP  
  (bring the book)
(iii) VP \rightarrow V NP that-S [i.e., CP]  
  (persuade John that there was no hope)
(iv) VP \rightarrow V PP  
  (decide on a new course of action)
etc.

Now among the many major lasting innovations in Aspects was the introduction of the lexicon into syntactic theory. (In the LSLT model, the source of words and grammatical morphemes was phrase structure rules.) Lexical entries included all that is particular and idiosyncratic about specific lexical items. Among that information was a subcategorization frame indicating for each verb, for example, exactly what kind of VP it can be inserted into. For example, the subcategorization frames for elapse and persuade were essentially as follows:

(5)  
elapse  [ ___ ]
persuade [ ___ NP CP]

After introducing this technology, Chomsky, commenting on (4), observes (p. 96) “Corresponding to each such string dominated by VP, there is a strict subcategorization of Verbs.” This is as things must be, of course, since what would it mean for a particular language to have a particular VP rule if that language had no verb that could ever be inserted into that VP? The converse is equally true. How could a particular language have a particular verb if there were no phrase structure rule to create the VP structure required by that verb? What Chomsky doesn’t point out is that there is thus total redundancy between subcategorization and PS rules at (what became) the V-bar level. This redundancy strongly suggests that one or the other should be eliminated. Further, since lexical information must be stated somewhere, it is PS rules that should go. At the time, this seemed improbable since phrase structure was so closely tied to phrase structure rules and phrase structure derivations. But eventually proposals began to appear for phrase structure without phrase structure rules or derivations. For example, McCawley (1968) attributes to Richard Stanley a theory of phrase structure in which the notion of ‘derivation’ is dispensed with: the base component is simply a set of node admissibility conditions. Lasnik and
Kupin (1977) offer a set-theoretic analogue of the Stanley/McCawley graph-theoretic proposal. For Lasnik and Kupin, any set of mono-strings (strings containing at most one non-terminal element) is a (reduced) phrase-marker, as long as it satisfies certain representational conditions on its completeness, consistency, etc. And just a bit later, in very influential work, Stowell (1981) argued extensively for the superfluity of phrase structure rules in the creation of phrase structure representations.

3 Linear Order in the Phrase Structure Component?

Chomsky observes in Aspects (p. 123) that the PS rules carry out two separate functions: They define hierarchical relations (crucial to the specification of grammatical relations), and they determine the linear ordering of elements in deep structure. He then suggests that “At least the first of these functions appears to be carried out in a very general and perhaps universal way by these rules.” Referencing Curry (1961) and Saumjan and Soboleva (1963), he then makes the following striking comment:

“It has been suggested several times that these two functions of the categorial component be more sharply separated, and that the second, perhaps, be eliminated completely.”

So, an old concern but also quite a modern one. However Chomsky’s position at the time was quite different from what it became in the 1990s. First, he summarizes the Curry and Saumjan-Soboleva position:

“They propose, in essence, that in place of such rules as (69), the categorial component should contain the corresponding rules (70), where the element on the right is a set rather than a string:

(69)  \[ S \rightarrow NP \rightarrow VP \]
\[ VP \rightarrow V \rightarrow NP \]

(70)  \[ S \rightarrow \{NP, VP\} \]
\[ VP \rightarrow \{V, NP\} \]

In (70), no order is assigned to the elements on the right-hand side of the rule; thus \( \{NP, VP\} = \{VP, NP\} \), although \( NP \rightarrow VP \neq VP \rightarrow NP \). The rules (70) can be used to define grammatical relations in exactly the way indicated for the rules (69). The rules (69) convey more information than the corresponding rules (70), since they not only define an abstract system of grammatical relations but also assign an abstract underlying order to the elements.” (p. 124)

Chomsky then reports that

“Proponents of set-systems such as (70) have argued that such systems are more ‘abstract’ than concatenation-systems such as (69), and can lead to a study of grammatical relations that is independent of order, this being a phenomenon that belongs only to surface structure.”
That sounds very modern indeed. But interestingly, Chomsky proceeds to strongly reject that argument:

“The greater abstractness of set-systems, so far as grammatical relations are concerned, is a myth. Thus the grammatical relations defined by (70) are neither more nor less ‘abstract’ or ‘order-independent’ than those defined by (69); in fact, the systems of grammatical relations defined in the two cases are identical. A priori, there is no way of determining which theory is correct; it is an entirely empirical question, and the evidence presently available is overwhelmingly in favor of concatenation-systems over set-systems, for the theory of the categorial component. In fact, no proponent of a set-system has given any indication of how the abstract underlying unordered structures are converted into actual strings with surface structures. Hence, the problem of giving empirical support to this theory has not yet been faced.” (pp. 124-125)

This was evidently true at the time, and for quite a while after, but decades later an intriguing proposal did appear precisely addressing the question of “how the abstract underlying unordered structures are converted into actual strings.” Based on the important insights of Kayne (1994), Chomsky (1995) argues that syntactic structures are, indeed, unordered and that the observed linear ordering arises via Chomsky’s version of Kayne’s Linear Correspondence Axiom, which in the mapping to PF maps asymmetric c-command into linear precedence. Following a tradition advanced within generative grammar by Reinhart (1976), Chomsky indicates that “There is no clear evidence that order plays a role at LF or in the [transformational] computation” (p. 334).

4 Conclusion

As with so many topics in syntactic theory, and with so many of his writings, Chomsky in Aspects set the agenda for generations of research. The lexicon, the syntactic cycle, the status of linear order are as important now as when they were first discussed. In all of these cases, the perspective has been altered, occasionally even reversed, by later developments and discoveries. But it is fair to say that it was the arguments and analyses of Aspects that made many of these later advances possible.

References


HOW TO TRIGGER ELEMENTS OF I-LANGUAGES

DAVID W. LIGHTFOOT
Georgetown University

1 Global Approaches to Acquisition

In *Syntactic structures* Chomsky viewed a grammar as a device to generate the sentences of a language (1957: 11). Linguists could establish what the grammar might be through a metric evaluating its success in generating a corpus in an optimal way rather than by following procedures to discover a grammar or to decide whether a grammar was correct or not for a given corpus (1957: ch. 6). In this respect Chomsky was reflecting the anti-positivist moves of the 1950’s (e.g. Popper 1959), showing that there is no predetermined path to scientific truth and that the best we can do is say which analyses are better for certain data. He argued against formal approaches that could not capture properties of natural languages (Markov processes and pure phrase structure grammars) and in favor of derivations relating different levels of abstraction (transformations).

*Syntactic structures* contains nothing about cognition, psychology, or the acquisition of language by children, nor does the long, influential review by Robert Lees (1957), apart from the mystified optimism of his final section on ‘learning theory.’ We know that as Chomsky wrote *Syntactic structures*, he was preparing to write his 1959 attack on Skinner’s behaviorist approach to analyzing the human language capacity and he was interacting with Eric Lenneberg, who was to pursue the biological approach to language earliest and furthest (Lenneberg 1967). Several years later, in the introduction to the published version of his dissertation, he wrote that he had viewed his approach to linguists justifying their grammars through an evaluation metric as having ‘a psychological analog’ in what a child does in acquiring a grammar in the first few years of life (1975: 11, 12, 36, etc.): children also use an evaluation metric, but subconsciously. However, he had viewed the psychological analog and discussion of cognition and biology as ‘too audacious’ for the time (1975: 35). For discussion, see my introduction to the 2002 second edition of *Syntactic structures*.

1 Many thanks for helpful comments on an earlier draft from some of my favorite readers: Cedric Boeckx, Elan Dresher, Tony Kroch, Anne Lobeck, Betty Tuller, and Marit Westergaard.
By 1965, either Chomsky had become more audacious or the field had shifted sufficiently to allow him to discuss the cognitive and biological consequences of *Syntactic structures*. Given that Chomsky had been thinking along these lines from the beginning, there was an easy translation of the methodological claims of the earliest work into substantive claims about human cognition and biology. The famous first chapter of *Aspects* lays it out carefully and makes matters of children’s acquisition central for linguists constructing their grammars. If grammars are learnable by children and meet other biological requirements, then there are great empirical demands that must be met, narrowing the range of viable hypotheses and thereby helping linguists to converge on plausible hypotheses and children to attain their mature system. In particular, we must show how children can “project” from the Primary Linguistic Data (PLD) that they experience to the mature grammars that they eventually attain, solving the profound “poverty-of-stimulus” problems. Chomsky viewed children as following an evaluation metric preferring certain grammars over others with respect to a given corpus of sentences.

Here I aim to tease apart the methodological goal of justifying analyses and the psychological goal of characterizing how children acquire their system and to suggest a different approach to childhood acquisition, which became viable after another major methodological move made by Chomsky twenty years after *Aspects*.

In *Aspects*, Chomsky wrote:

> Certain problems of linguistic theory have been formulated as questions about the construction of a hypothetical language-acquisition device ... We may think of the theorist as given an empirical pairing of collections of primary linguistic data associated with grammars that are constructed by the device on the basis of such data. Much information can be obtained about both the primary data that constitute the input and the grammar that is the “output” of such a device, and the theorist has the problem of determining the intrinsic properties of a device capable of mediating this input-output relation. (Chomsky 1965: 47)

Chomsky writes here of ‘collections of PLD’ being ‘associated with grammars’ and a common approach to children evaluating their grammars has been **GLOBAL**: children evaluate postulated grammars as wholes against the whole corpus of PLD that a child encounters, checking which grammars generate which data.

For example, Robin Clark’s genetic algorithm (1992) employs his Fitness Metric to measure very precisely the fitness of grammars in generating sets of sentences. The key idea is that certain grammars yield an understanding of certain sentences and not others; put differently, they generate certain sentences and not others. The Fitness Metric quantifies the failure of grammars to parse sentences by counting the “violations,” sentences experienced that cannot be generated by the grammar being evaluated. There are two other factors involved in his Fitness equation, a superset penalty and an elegance measure, but those factors are subject to a scaling condition and play a minor role, which I ignore here. The Fitness Metric remains the most sophisticated and fully worked out evaluation measure that I know. It is a global measure of success, assigning indices to whole, fully formed grammars against a whole corpus of sentences.²

---

² It is worth noting that that the metric is technically and conceptually flawed insofar as it is based on an assumption that grammars with a greater number of parameters set correctly will be the fittest, the most successful in parsing/generating incoming data but Dresher (1999: 54-58) demonstrated that this assumption is false, that there is
Gibson & Wexler (1994) take a different approach but their children also effectively evaluate whole grammars against whole sets of sentences, although they react to local particularities; their children are ‘error-driven.’ They acquire a mature grammar eventually by using a hypothetical grammar (a collection of parameter settings) and revising it when they encounter a sentence that their current grammar cannot generate, an “error.” In that event, children follow the Triggering Learning Algorithm to pick another parameter setting and continue until they converge on a grammar for which there are no unparsable PLD, no errors. Gibson & Wexler used a toy system of three binary parameters that define eight possible grammars, each of which generates a set of sentence types. The child converges on a whole grammar in light of the complete set of sentence types experienced, essentially testing the generative capacity of grammars against a corpus of sentences.

There are major feasibility problems for both these global, evaluation-based approaches to language acquisition, evaluating whole grammars against comprehensive sets of sentences experienced (Lightfoot 2006: 76f). If there are thirty points of binary choices (a conservative estimate), there are over a billion grammars to evaluate against the data set; if there are forty points of variation, then over a trillion grammars to evaluate; if there are fifty points of variation, then the numbers become astronomical. And bear in mind that each grammar generates an infinite set of sentences and structural descriptions. In addition, in order to check whether the generative capacity of a grammar matches what the child has heard, s/he will need to retain a memory in some fashion of everything that has been heard. Such an approach to language acquisition faces great difficulties when one thinks beyond toy systems with just a few parameter settings.

Alongside the feasibility issues, one’s theory must accommodate change in I-languages from one generation to another. Seeking to explain changes through language acquisition requires a different approach to language acquisition. If children attain grammars by evaluating them against a corpus of data, then they would need to be confronted with the data generated by the new grammar in order to first select the grammar that generates them. This introduces problems of circularity: what comes first, the new grammar to generate the new data or new data that require the child to select the new grammar?

2 Children Discovering Elements of I-languages

A different approach would separate the methodological and psychological components. One would acknowledge the validity of Chomsky’s 1957 arguments that linguists need to evaluate the success of their hypothesized grammars rather than to seek a discovery procedure leading to correct hypotheses or a means to decide whether a hypothesis is correct or not. However, one might view children differently, acquiring their mature grammar by following a kind of subconscious discovery procedure, albeit not the kind of discovery procedure contemplated by Chomsky (1957) but one incorporating a rich theory of Universal Grammar (UG).

In the early work, Chomsky had taken children to be acquiring grammars that generated socially defined languages like English, Turkish, or Warlpiri and he postulated ‘an ideal speaker-

---

no smooth correlation between accuracy in what is parsed/generated and the number of parameters set correctly. See Lightfoot (2006: 74) for discussion.
listener, in a completely homogeneous speech-community, who knows its language perfectly’ (1965: 3), idealizing away from individual variation.

In 1986, he made a major methodological move, abandoning the idealized speaker-listener and the idea of socially defined languages as having any kind of psychological reality. He embraced language variation and distinguished between external E-language and internal, individual, intensional I-languages as the proper focus of linguistic theory. In effect, he adopted Wilhelm von Humboldt’s (1836) distinction between the languages of nations and those of individuals. Hermann Paul (1877: 325) emphasized the individual and biological view of language, noting in an early work ‘dass die reelle Sprache nur im Individuum existiert’ (‘real language exists only in individuals’). Later, Paul (1880: 31) attacked the group psychology of Lazarus and Steinthal and wrote that ‘Wir müssen eigentlich so viele Sprachen unterscheiden als es Individuen gibt’ (‘we must in fact distinguish as many languages as there are individuals’).

Chomsky (1986) followed von Humboldt and Paul and distinguished external E-language and internal, individual I-languages. E-language refers to language out there in the world, the kind of thing that a child might be exposed to, an amorphous, mass concept. I-language, on the other hand, refers to a biological system that grows in a child’s mind/brain in the first few years of life and characterizes that individual’s linguistic capacity. It consists of structures, categories, morphemes, phonemes, and features, and computational operations that copy items, delete them, assign indices to them, etc. One’s I-language is a private object that permits thought and partial communication with certain other speakers, although it may differ from the I-languages of those speakers.

Everything was now individualized and linguists were free to postulate different systems for different individuals, allowing the possibility of billions of grammars or, now, “I-languages.” Indeed, a little later, Kroch (1989) postulated the idea that individuals may use multiple co-existing grammars, which turned out to be very productive for work on syntactic change and opened the possibility of billions and billions of grammars, which could be studied through the methods of population biology.

These ideas of E-language and I-languages suffice for the purposes of accounting for language acquisition and we do not need the conventional, socially defined English or Estonian. One can view children as acquiring their individual I-language on exposure to external E-language. Rather than evaluating systems against a set of data, children can be viewed as paying no attention to what any I-language or grammar generates but instead growing an I-language by identifying and acquiring its elements (Lightfoot 2006). Children parse the E-language they hear and discover the categories and structures needed to understand what they hear, thereby accumulating the elements of their I-language (for this “cue-based acquisition,” see Dresher 1999, Lightfoot 1999, and Sakas & Fodor 2001).3

UG provides children with the structures that they might need in order to understand and parse their ambient E-language; robust E-language triggers them. Children are born to parse and at a certain stage of development, after they know that cat is a noun referring to a domestic feline and sit is an intransitive verb, they may hear an expression The cat sat on the mat and recognize that it contains a Determiner Phrase (DP) consisting of a determiner the and a noun cat and a Verb Phrase (VP) containing an inflected verb sat (V+I) followed by a Preposition Phrase (PP) on the mat, i.e. vp[V+I PP]. The child makes use of the structures needed to parse what is heard (i.e. the structures “expressed” by the E-language experienced) and, once a structure is used, it is incorporated into the emerging I-language. In this way, a child discovers and accumulates the

---

3 The “cues” terminology has given rise to confusions and misunderstanding and I drop it here.
elements of its I-language, which are required to parse the E-language around them; children acquire elements of their I-language in piecemeal fashion.

At no stage does the child calculate what its current I-language can generate; rather s/he simply accumulates the necessary structures and the resulting I-language generates what it generates. Furthermore, if UG makes available a thousand possible structures for children to draw from, that raises no intractable feasibility problems comparable to those facing a child evaluating the generative capacity of grammars with thirty possible parameter settings, checking the grammars against what has been heard. It involves no elaborate calculations. Children developing some form of English I-language learn without apparent difficulty irregular past tense and plural forms for a few hundred verbs and nouns. Learning that there is a structure $VP[V+I\ PP]$ (see previous paragraph) seems to be broadly a similar kind of learning, similar to acquiring the irregular past tense forms, although much remains to be said.\(^4\)

Under this approach to language acquisition, children pay no attention to the generative capacity of their emerging I-language but they attain particular elements of it step-by-step (see Dresher 1999 on the resulting ‘learning path’) and they do that by experiencing particular elements of E-language that they must attempt to parse. Work seeking to explain diachronic changes through acquisition has enabled us to link changes in E-language to changes in particular elements of I-languages, giving us a clear idea of what triggers what in some cases. I will illustrate this by sketching briefly a phase transition in the history of English that is now well understood.

### 3 Identifying Triggers

There is good reason to believe that English I-languages underwent a change whereby words like *can, could, must, may, might, will, would, shall, should* and *do* were once categorized as more or less normal verbs but then were re-categorized as Inflection elements in all known grammars of English speakers after the time of Sir Thomas More in the early sixteenth century. For More and his forebears, verbs like *can* moved to a higher Inflection position, as in (1), and after More they were generated directly as Inflection elements and occurred in structures like (2), a single shift in the system, which was manifested by the simultaneous loss of several phenomena, the phase transition. Early speakers had used expressions like *He has could see stars, Canning see stars ..., She wanted to can see stars, She will can see stars, and She can music.* Such sentences may be generated by an I-language with structures like (1), and not by I-languages with structures like (2). The SINGULARITY of the change in I-languages explains the PARALLELISM in the loss of these phenomena (for details, see Lightfoot 2006, in press a, b).

---

\(^4\) These structures may affect many sentences or only a few. Tony Kroch notes that in modern English we have remnants of verb-second in *Kim left and so did Jim*. Such examples look like irregular verbs, residues of once productive processes.
A critical property of this change is that it consisted entirely in the loss of phenomena and there were no new forms emerging. Since children converge on their I-language in response to ambient simple expressions and not in response to “negative data” about what does not occur, the new, less diverse data need to be explained by a new abstract system that fails to generate the obsolescent phenomena. There were no new forms in which the modal auxiliaries began to occur and so the trigger for the new system must lie elsewhere. In this case, the new PLD cannot be the new output of the new grammars, not only because that would be circular but because, in any case, there are no new forms. Changes like this, which consist only in the loss of expressions, make up a kind of poverty-of-stimulus argument for diachrony: there appear to be no new forms in the PLD that directly trigger the loss of those expressions.

If we ask why this or any other I-language change happened, there can only be one answer under an approach explaining changes through acquisition: children came to have different PLD as a result of a prior change in external language. We have a good hypothesis about what the prior E-language change was in this case.

Early English had complex morphological properties. For example, we find *fremme, fremst, fremþ, fremmoþ* in the present tense and *fremed, fremedest, fremede, fremedon* in the past tense of ‘do;’ *sêo, siehst, siehþ, sêop* in the present tense for ‘see;’ *rîde, rîst, rîtt, rîdoþ* for the present tense of ‘ride,’ and *råd, ride, råd and ridon* for the past tense. There was a massive loss of verb morphology in Middle English, beginning in the north of England and due to intimate contact with Scandinavian speakers, leaving the third person singular –s ending as the only surviving element of present tense verb morphology. Again I skip interesting details but the crucial factor is that external language that children heard changed such that the modern modal auxiliaries like *can, shall*, etc came to be morphologically distinct from other verbs; as members of the small preterite-present class, they always lacked the one surviving feature of present tense verb morphology, the –s ending of the third person singular. Furthermore, their “past tense” forms (*could, would, might*, etc) had meanings that were not past time, reflecting old subjunctive uses (3). The evidence indicates that these modal verbs were re-categorized in people’s internal
systems, because they had become formally distinct from other verbs as a result of the radical simplification of morphology. So we see domino effects: changes in what children heard, the newly reduced verb morphology, led to a different categorization of certain verbs, which yielded systems (2) that were not compatible with the obsolescent data.

(3) They might/could/would leave tomorrow.

More was the last known speaker with the old system. For a period, both systems co-existed: some speakers had (1) and others had (2), the former becoming rarer over time, the latter more numerous. A large literature is now devoted to this kind of sociological variation, changing over time, and the spread of the new system can be studied through the methods of population biology (Lightfoot in press, b). As some speakers began to have new I-languages, that affected the ambient E-language, making it more likely that members of a speech community would acquire the new I-language. Partha Niyogi (2006) provided interesting computational simulations of this approach to language acquisition and diachronic change.

An old idea holds that one can sometimes discover properties of systems by seeing how they change. The major contributions of diachronic work in syntax lie in explaining one kind of variation, due to co-existing I-languages, and in revealing what the E-language trigger might be for any particular property of I-languages. In particular, by looking at change one can sometimes generate productive hypotheses about which PLD trigger which elements of I-languages, associating particular PLD drawn from E-language with particular elements of emerging I-languages. From a certain perspective, one may paraphrase Theodosius Dobzhansky: nothing in (language-specific) syntax makes sense except in the context of change – that is, ideally for any language-specific syntactic property, one would show how it might have emerged, just as ‘nothing in biology makes sense except in the light of evolution’ (Dobzhansky 1973).

It is important to recognize that the same elements of I-language may be triggered by different PLD in different languages. For example, we have seen that the new Inflection category in English I-languages was triggered by new properties of verb morphology. O’Shannessy (2013) reports a new language also with an Inflection category but triggered by quite different features of E-language having to do with irrealis properties. Similarly, children developing an English I-language could learn that VP’s have V-complement order from a simple sentence like Bent visited Oslo; since verbs do not move in English I-languages and Inflection lowers on to verbs, visited Oslo can only be analyzed as \( v_p[\text{visit}+I \text{Oslo}] \). A Norwegian child, however, could not draw the same conclusion from the word-for-word translation Bent besøkte Oslo, which does not reveal the underlying position of the verb. Norwegian I-languages are verb-second: finite verbs move to a high, “second” position (presumably in the C projection), yielding simple structures like \( cp[\text{Bent cbesøkte v}_p[\text{besøkte Oslo}]] \). The verb-second analysis is required by synonymous sentences like Oslo besøkte Bent, where the finite verb surfaces to the left of the subject DP and requires a structure \( cp[\text{Oslo cbesøkte v}_p[\text{Bent v}_p[\text{besøkte Oslo}]]] \), as well as “topicalized” expressions like \( cp[\text{Søndager cbesøkte v}_p[\text{Bent v}_p[\text{besøkte Oslo}]]] \) ‘On Sundays Bent visited Oslo.’ Therefore, Bent besøkte Oslo does not reveal the structure of the VP and a more complex sentence like Bent kann Oslo besuchen ‘Bent can visit Oslo’ is needed to express the \( v_p[V \text{DP}] \) structure, Bent kann v_p[besøke Oslo]. Similarly, the complex Bent kann Oslo besuchen

---

5 The syntactic evidence for the verbal status of the modals is not directly affected by the new morphology. Therefore children must have ignored some of the input data if they re-categorized the modals; those data simply became obsolete.
reveals the \( VP[DP \ V] \) structure of German, another verb-second language but with complement-verb order.

### 4 Minimalism and Acquisition

Lightfoot (1999) postulated that UG provide the structures \( V, VP[V \ DP], VP[DP \ V], \) and \( \text{SpecCP}[XP] \), available to children if their PLD triggered them. In addition, Lightfoot (2006: 82-86) hypothesized the eleven structures of (4) to capture a wide range of properties involving various syntactic structures, results of computational operations, binding relations, and intonation patterns. They presupposed a substantial contribution from UG (including a condition on deletion, a binding theory, and a bottom-up analysis of intonation structure, and more) and they were each triggered by the sentences on the right, which expressed or required the relevant structure (for details, see Lightfoot 2006).

(4)

- a. Ce
- b. CP[wh-
- c. ve
- d. NP+I
- e. \( VP[DP \ # DP] \)
- f. DP[DP \ D \ NP]  
  - g. \( \text{SpecCP}[DP \ # \ V \ DP[-self_1]] \)
  - h. \( \text{SpecCP}[DP \ # \ V \ DP[-self_1]] \)
  - i. \( \text{SpecCP}[DP \ # \ V \ DP[-self_1]] \)
  - j. \( \text{SpecCP}[DP \ # \ V \ DP[-self_1]] \)
  - k. \( \text{SpecCP}[DP \ # \ V \ DP[-self_1]] \)

Postulating such UG structures enabled us to solve a wide range of poverty-of-stimulus problems. The structures constituted rich and structurally complex information that would have evolved through genetic mutation. That is part of a tradition noted by Chomsky (2007: 2): ‘At the time of the 1974 discussions, it seemed that [the faculty of language] must be rich, highly structured, and substantially unique.’ There is nothing inherently untenable about that, given that genetic mutations hold at a biochemical level and the structures at a cognitive level and that we know next to nothing about how such levels comport. Indeed, if (4) is what is required to solve poverty-of-stimulus problems, then one celebrates the explanations achieved.

However, one always seeks better alternatives. Since *Aspects*, work in generative syntax has gone through a number of distinct paradigms. Beginning with *Aspects* and flourishing through the 1970’s we had conditions on the expressive power of rules, then a model where principles of Government and Binding were central, followed soon by the Principles-and-Parameters approach focusing on language variation and on how the variation might be attained by children exposed only to simple, primary data. In fact, considerations of acquisition were central from the time of *Aspects* through the Principles-and-Parameters period.

Now we are in a paradigm of Minimalism, which is driven by quite different forces, seeking to minimize the information of the linguistic genotype and seeking simpler alternatives to the structures of (4). There is little substantive discussion of variation and even less of acquisition and, as Boeckx (2011: 207) puts it, ‘central Minimalist tenets clash with the traditional
Throughout the modern history of generative grammar, the problem of determining the character of [the faculty of language] has been approached “from the top down”: How much must be attributed to UG to account for language acquisition? The [Minimalist Program] seeks to approach the problem “from the bottom up”: How little can be attributed to UG while still accounting for the variety of I-languages attained?

The goal of Minimalism is to see what information that has been apparently needed at UG can be explained in terms of more general principles or be shown to be unnecessary.

Proposals within the Minimalist Program provide severe, general, and sometimes elegant limits on the structures available for narrow syntax. Boeckx (2011: 210) advocates a ‘Strong Uniformity Thesis,’ whereby ‘principles of narrow syntax are not subject to parameterization’ and variation ‘would be confined to the margins of narrow syntax, especially the morphophonological component (PF).’ Along those lines, Gianollo, Guardiano, & Longobardi (2008) took variation seriously and explored the possibility of a theory of syntactic variation beyond narrow syntax in the properties of functional features, which may be grammaticalized, checked by a lexical category, ‘spread on’ a lexical category, or checked by a strong feature. And, of course, there are many other proposals. By embracing the radical limits imposed by current Minimalist work, perhaps we could find how those limited structures are expressed in what children hear and then determine which structures are expressed robustly enough to be incorporated into children’s I-languages. We would abandon evaluation metrics and checking the generative capacity of I-languages against what is observed and instead seek a means by which children may DISCOVER the structural elements that are EXPRESSED (i.e. required) by the E-language that they are exposed to, tweaking Aspects’ global approach to acquisition to the local approach advocated here. The research agenda would be to seek to establish what particular aspects of PLD might trigger particular structural elements hypothesized by Minimalist work; the challenge would be to show that there are plausible PLD that could have the necessary triggering effects. In this way we could seek to resolve Boeckx’s ‘clash’ between Minimalist tenets and the traditional concern with acquisition.

In a similar spirit, Yang (2010) argues for the complementarity of structures provided by UG and statistical generalization. Structures need to be expressed by E-language robustly. Lightfoot (1999) sought to explore some effects in changes in robustness of expressions triggering elements of I-languages but it is important to remember that there is unlikely to be a uniform definition of robustness: some I-language elements are expressed with some frequency and others may be triggered by single events (such as the learning of certain vocabulary items).

I have argued here that one can sometimes identify aspects of E-language that trigger specific elements of I-languages. When such triggers can be identified, they may cast light on how the changing distribution of E-language properties trigger structural elements of new I-languages. Rather than abandoning any concern with explaining language acquisition as if it were too mysterious to explain, we might link the “local” approach to acquisition, seeking E-language triggers for specific elements of I-languages, with the range of structures allowed by our favorite Minimalist proposals. What is NOT a rational approach is to assert that pre-Minimalist ideas about UG are too complex to have evolved through genetic mutation and therefore must be discarded. Given the limitations on our understanding of the evolution of the language faculty (Hauser et al. 2014) and the richness of what we have learned about the acquisition of language,
that would be foolish until we have some kind of alternative offered.

References


THE GARDEN OF EDEN PERIOD FOR DEEP STRUCTURE AND SEMANTICS

BARBARA H. PARTEE
University of Massachusetts, Amherst

1 Syntax and Semantics before Aspects

At the time of Aspects (Chomsky, 1965), the expression “syntax-semantics interface” was still decades away from its first appearance, but questions about the potential relation between syntax and semantics had figured in Chomsky’s writings from the start. Chomsky clearly considered semantics important in principle, while expressing considerable skepticism about the prospects for developing a scientific theory of meaning. Between Syntactic Structures and Aspects, work by Katz & Fodor and Katz & Postal influenced Chomsky to become cautiously optimistic about the possibility of a semantic component for generative grammar. The theory presented in Aspects was characterized by an elegant architecture in which deep structure formed the input to semantic interpretation, transformations mapped deep structure to surface structure, and that in turn formed the input to phonological interpretation. The brief period when the Aspects theory held sway is what I like to call the “Garden of Eden period”. I conclude with a brief review of some of the causes of the “expulsion from the Garden” that soon followed.

1.1 Before Syntactic Structures

In the century before Syntactic Structures, linguistics became at least in part a science, but semantics was not part of that development. Linguistics began to emerge as a science in the nineteenth century with the German Junggrammatiker’s breakthrough discoveries about sound changes and the evolutionary history of the Indo-European languages. Darwin in his Origin of Species “said that linguistics, as practiced by the leading exponents of comparative Indo-European philology, offers the paradigm of scientific method” (Harris and Taylor, 1997, p.187).

---

1 I say nothing in this note about the explosion of work in semantics and the syntax-semantics interface that began soon after Aspects, or about semantics in other linguistic traditions or in philosophy and logic. See Partee 2011, 2014; all of these issues will be discussed in Partee in preparation.
And in both Europe and the U.S., there was a self-conscious drive to view linguistics as a science in the 1930’s; linguistics was part of the Vienna Circle’s “unified science” movement, with the semiotician Charles Morris as one of its leaders.

As for semantics, there was a mixture of negative attitudes and neglect in American linguistics in the 20th century. There had been rather little semantics in early American anthropological linguistics, since linguistic fieldwork generally had to start with phonetics, then phonology, then morphology, then perhaps a little syntax; any work on semantics was limited to dictionary-making and semantic features for the structural analysis of kinship terms and other such lexical domains. The behaviorists viewed meaning as an unobservable aspect of language, not fit for scientific study, and that had some influence on the Bloomfieldians. Quine had strong philosophical skepticism about the concept of meaning, and his and Goodman’s influence on Chomsky is acknowledged in the preface to *Syntactic Structures*.

Chomsky has frequently expressed ambivalence about semantics as a science and about the relation between syntax and semantics. This is evident even before *Syntactic Structures* in *The Logical Structure of Linguistic Theory* (LSLT; Chomsky, 1975b). In *LSLT*, Chomsky writes in the first chapter:

> I will merely emphasize again that the “legitimacy of semantics” (whatever that might mean) has never, to my knowledge, been challenged, nor has there ever been any question of the importance of incorporating a study of reference, meaning, language use, and related topics within a full theory of language that will deal, in particular, with the highly significant relations between formal structure and semantic interpretation. The appeal to meaning must be clearly distinguished from the study of meaning. The latter enterprise is unquestionably central to the general theory of language, and a major goal of the *SS-LSLT* approach is to advance it by showing how a sufficiently rich theory of linguistic form can provide structural descriptions that provide the basis for the fruitful investigation of semantic questions. (Chomsky, 1975b, 21)

But also in *LSLT* he writes “Part of the difficulty with the theory of meaning is that “meaning” tends to be used as a catch-all term to include every aspect of language that we know very little about” (Chomsky, 1975b, fn 21, p. 97).

So before *Syntactic Structures* (Chomsky, 1957) there was no syntax-semantics interface to think about in linguistics.

### 1.2 In *Syntactic Structures*: Chomsky on Semantics in 1957

*Syntactic Structures* contains clear evidence of Chomsky’s ambivalence about semantics, with assertions of its importance along with insights into challenges facing any attempt to construct a systematic theory relating syntax and semantics.

He spends many pages arguing that semantic notions are of no use in constructing a grammar, and arguing that intuitions of grammaticalness are distinct from intuitions of meaningfulness. “Grammar is best formulated as a self-contained study independent of semantics. In particular, the notion of grammaticalness cannot be identified with meaningfulness.” (p.106)
But at the same time he holds that one test of a good syntax is that it should provide a good basis for a good semantics (if we had any idea how to study semantics). “In other words, we should like the syntactic framework of the language that is isolated and exhibited by the grammar to be able to support semantic description, and we shall naturally rate more highly a theory of formal structure that leads to grammars that meet this requirement more fully.” (p.102)

And he argues that transformational grammar is a positive step in that direction, since it uncovers differences at the “transformational level” that are obscured in the output. “The general problem of analyzing the process of ‘understanding’ is thus reduced, in a sense, to the problem of explaining how kernel sentences are understood, these being considered the basic ‘content elements’ from which the usual, more complex sentences of real life are formed by transformational development.” (p.92).

In Syntactic Structures, he illustrates this point with an example of ambiguity that cannot be captured at the level of phrase structure: he argues for a transformational derivation of the ambiguous (1a) from two different kernel sentences, (1b) and (1c) (pp.88-89). He takes this as an example of a property that grammars should have: a correspondence between constructional ambiguity and “real cases of ambiguity” (p.86).

(1) a. the shooting of the hunters
    b. The hunters shoot.
    c. They shoot the hunters.

But Chomsky also notes that transformations sometimes change meaning. “… we can describe circumstances in which a ‘quantificational’ sentence such as [(2a)] may be true, while the corresponding passive [(2b)] is false, under the normal interpretation of these sentences—e.g., if one person in the room knows only French and German, and another only Spanish and Italian. This indicates that not even the weakest semantic relation (factual equivalence) holds in general between active and passive.” (pp. 100–101)

(2) a. Everyone in this room speaks two languages.
    b. Two languages are spoken by everyone in this room.

In later years, those judgments about (2) came to be questioned; it has been claimed either that both (2a) and (2b) are ambiguous, or that (2a) is unambiguous but the passive (2b) is ambiguous. Chomsky himself has a long footnote in Aspects noting that it is possible that both sentences are ambiguous, but with different readings preferred for (2a) and (2b).

His general worries about the study of meaning are clearly expressed in Syntactic Structures:

In proposing that syntactic structure can provide a certain insight into problems of meaning and understanding we have entered onto dangerous ground. There is no aspect of linguistic study more subject to confusion and more in need of clear and careful formulation than that which deals with the points of connection between syntax and semantics. The real question that should be asked is: “How are the syntactic devices available in a given language put to work in the actual use of this language?” (p. 93)

---

2 This is footnote 9 to Chapter 3 of Aspects, pp. 224-225. We discuss it in Section 2.2 below.
I would sum up Chomsky’s attitude toward semantics in 1957 approximately as follows: Semantics is in principle an important part of linguistic theory; a theory of syntax should probably be judged in part by the support it provides for semantic interpretation; but semantics itself is a subject not so far amenable to scientific investigation, and it is misguided to appeal to meaning in the explanation of syntactic phenomena. The seriousness with which Chomsky took semantics may be seen in the final paragraph of *Syntactic Structures*, which ends with a cautious call for the study of syntax and semantics and their interface:

More generally, it appears that the notion of “understanding a sentence” must be partially analyzed in grammatical terms. To understand a sentence it is necessary (though not, of course, sufficient) to reconstruct its representation on each level, including the transformational level where the kernel sentences underlying a given sentence can be thought of, in a sense, as the ‘elementary content elements’ out of which this sentence is constructed. In other words, one result of the formal study of grammatical structure is that a syntactic framework is brought to light which can support semantic analysis. Description of meaning can profitably refer to this underlying syntactic framework, although systematic semantic considerations are apparently not helpful in determining it in the first place. The notion of “structural meaning” as opposed to “lexical meaning”, however, appears to be quite suspect, and it is questionable that the grammatical devices available in language are used consistently enough so that meaning can be assigned to them directly. Nevertheless, we do find many important correlations, quite naturally, between syntactic structure and meaning; or, to put it differently, we find that grammatical devices are used quite systematically. These correlations could form part of the subject matter for a more general theory of language concerned with syntax and semantics and their points of connection. (pp. 107–108)

1.3 Between *Syntactic Structures* and *Aspects*: Katz, Fodor, Postal

Jerrold Katz and Jerry Fodor were the first to work on adding a semantic component to generative grammar (Fodor, 1961, Katz, 1961, Katz and Fodor, 1962, 1963). They were concerned with compositionality, which they generally called the Projection Problem: how to get the meaning of a sentence from meanings of its parts. Hodges (1998) identifies their 1963 paper as the first use of the term “compositionality”. “As a rule, the meaning of a word is a compositional function of the meanings of its parts, and we would like to be able to capture this compositionality” (p. 501 in the version reprinted in (Fodor and Katz, 1964)). “…the fact that a speaker can understand any sentence must mean that the way he understands sentences he has never previously encountered is compositional: on the basis of his knowledge of the grammatical properties and the meanings of the morphemes of the language, the rules the speaker knows enable him to determine the meaning of a novel sentence in terms of the manner in which the parts of the sentence are composed to form the whole.” (p. 482 in the version reprinted in Fodor and Katz 1964).

Since at that time, “Negation” and “Question Formation” were transformations of affirmative and declarative sentences, transformations could clearly change meanings, and meaning depended on the entire transformational history. Katz and Fodor’s semantic Projection rules applied to T-markers, which were extensions of the phrase structure P-markers.
Katz and Fodor took compositionality seriously at the outset of their work on semantics; theirs is probably the first theory of the “syntax-semantics interface”. Their semantic tools were quite primitive, however, with semantic representations consisting of sets of “semantic features”, undoubtedly influenced by feature theories of phonology and componential analysis of kinship terms, etc., and quite unsuited for anything beyond the decomposition of one-place predicates. The idea that truth conditions might be relevant to semantics did not penetrate linguistics until after 1970.

In a theoretically important move related to the problem of compositionality, Katz & Postal (1964) made the innovation of putting such morphemes as Neg and a question morpheme Q into underlying structure, arguing that there was independent syntactic motivation for doing so. Then meaning could be determined on the basis of underlying structure alone. The Katz-Postal hypothesis was an interesting and provocative claim, and even without any “real semantics” at the foundation, it led to interesting debates about apparent counterexamples, which we return to in Section 3. And it became one of the pillars of the Aspects theory.

# 2 Semantics in Aspects

Among the innovations in Chomsky’s Aspects, the two that are central to the development of semantics are the syntactic innovations that led to “deep structure”, and the incorporation of the Katz-Postal hypothesis that transformations preserve meaning. Chomsky’s adoption of Katz and Postal’s conception of semantics is explicit in Footnote 10 to Chapter 1, Methodological Preliminaries: “Aside from terminology, I follow here the exposition in Katz and Postal (1964). In particular, I shall assume throughout that the semantic component is essentially as they describe it …” (p.198)

## 2.1 Deep Structure and the Katz-Postal Hypothesis

Where “generalized transformations” had previously been responsible for assembling kernel sentences into multiclausal structures, in Aspects the base phrase structure component is responsible for generating deep structures that assemble all the parts in their underlying forms. That structure together with the Katz-Postal hypothesis made it possible for Chomsky in Aspects to propose that deep structure is the input to semantics.

Thus the syntactic component consists of a base that generates deep structures and a transformational part that maps them into surface structures. The deep structure of a sentence is submitted to the semantic component for semantic interpretation, and its surface structure enters the phonological component and undergoes phonetic interpretation. The final effect of a grammar, then, is to relate a semantic interpretation to a phonetic interpretation—that is, to state how a sentence is interpreted. This relation is mediated by the syntactic component of the grammar, which constitutes its sole “creative” part. (Chomsky 1965, pp. 135–136)

---

2.2 Chomsky on Semantics in *Aspects*

The basic architecture of the theory presented in *Aspects* was elegant and attractive. And probably not coincidentally, some of Chomsky’s most positive statements about semantics occur in this period. The following paragraph in Chapter 3 not only underscores the idea of deep structure as the input to semantics, but asserts that such an idea had “motivated the theory of transformational grammar since its inception.”

Thus when we define "deep structures" as "structures generated by the base component," we are, in effect, assuming that the semantic interpretation of a sentence depends only on its lexical items and the grammatical functions and relations represented in the underlying structures in which they appear\(^4\). This is the basic idea that has motivated the theory of transformational grammar since its inception (cf. note 33, Chapter 2). Its first relatively clear formulation is in Katz and Fodor (1963), and an improved version is given in Katz and Postal (1964), in terms of the modification of syntactic theory proposed there and briefly discussed earlier. The formulation just suggested [with deep structures generated by the base component] sharpens this idea still further. In fact, it permits a further simplification of the theory of semantic interpretation presented in Katz and Postal (1964), since Transformation-markers and generalized transformations, as well as "projection rules" to deal with them, need no longer be considered at all. This formulation seems to be a natural extension and summary of the developments of the past few years that have just been summarized.

Notice that in this view one major function of the transformational rules is to convert an abstract deep structure that expresses the content of a sentence into a fairly concrete surface structure that indicates its form. (p.136)

And a little later in the same chapter we find an optimistic statement about compositionality and the possibility that the compositional interpretation rules might be universal:

In a somewhat similar way, the projection rules of the semantic component operate on the deep structure generated by the base, assigning a semantic interpretation (a "reading") to each constituent, on the basis of the readings assigned to its parts (ultimately, the intrinsic semantic properties of the formatives) and the categories and grammatical relations represented in the deep structure. (See Katz and Fodor, 1963; Katz and Postal, 1964; and other papers by Katz listed in the bibliography.) To the extent that grammatical categories and relations can be described in language-independent terms, one may hope to find universal projection rules, which need not, therefore, be stated as part of a specific grammar. (p. 144)

Looking back at these endorsements of the goal of developing a compositional semantics, I can see why I was so surprised by Chomsky’s later negative reactions to early attempts to combine Montague’s compositional formal semantics with Chomsky’s transformational syntax (Chomsky, 1975a). And Chomsky’s remarks later in Chapter 3 about semantic puzzles not

\(^4\) At this point in the text there appears Footnote 9, which contains a discussion of the problem posed by sentences with quantifiers like (2a-b) above, and to which we return at the end of this subsection.
solved by available means could be seen as potential encouragement of the direction taken soon after by generative semanticists, who may also have been surprised at his reaction to their work.

Finally, it is important to be aware of the many other problems that face a theory of semantic interpretation of the kind referred to in the preceding discussion. It is clear, as Katz and Fodor have emphasized, that the meaning of a sentence is based on the meaning of its elementary parts and the manner of their combination. It is also clear that the manner of combination provided by the surface (immediate constituent) structure is in general almost totally irrelevant to semantic interpretation, whereas the grammatical relations expressed in the abstract deep structure are, in many cases, just those that determine the meaning of the sentence. Cf., for example, Chapter 1, §4, and Ch 2, § 2.2. However, there are cases that suggest the need for an even more abstract notion of grammatical function and grammatical relation than any that has been developed so far, in any systematic way. Consider, for example, these sentence pairs:

(109) (i) John strikes me as pompous - I regard John as pompous
(ii) I liked the play - the play pleased me
(iii) John bought the book from Bill - Bill sold the book to John
(iv) John struck Bill - Bill received a blow at the hands of John

Clearly, there is a meaning relation, approaching a variety of paraphrase, in these cases. It is not expressible in transformational terms, as is possible, for example, in these cases:

(110) (i) John is easy for us to please - it is easy for us to please John

… Consequently, it seems that beyond the notions of surface structure (such as "grammatical subject") and deep structure (such as "logical subject"), there is some still more abstract notion of "semantic function" still unexplained. Various formal devices for expressing these facts suggest themselves, but the general problem seems to me non-trivial. (pp. 161-2)

We noted at the beginning of this subsection that when Chomsky first states his assumption that semantic interpretation depends only on underlying structure, he adds a footnote about the problems of some sentences with quantifiers. In particular, he is concerned there with the same active-passive pair (2a-b) that he used in *Syntactic Structures* as evidence that transformations sometimes change meaning. In *Aspects*, the discussion is much more nuanced, with the possibility that pragmatic and other factors could explain a difference in preferred interpretation, and perhaps the semantics could correctly assign both scope interpretations to both sentences.

Footnote 9: As it stands, this claim seems to me somewhat too strong, though it is true in one important sense of semantic interpretation. For example, it seems clear that the order of “quantifiers” in surface structures sometimes plays a role in semantic interpretation.

---

5 The examples in (109) suggest a need for a more abstract grammatical structure only if sentences which are (near-)semantic paraphrases must have a common (or very similar) deep structure. But compositionality does not impose such a requirement: with a richer semantics, sentences with quite different syntactic structures can be assigned the same or similar meanings.
Thus for many speakers – in particular, for me – the sentences “everyone in this room knows at least two languages” and “at least two languages are known by everyone in this room” are not synonymous. Still, we might maintain that in such examples both interpretations are latent (as would be indicated by the identity of the deep structures of the two sentences in all respects relevant to semantic interpretation), and that the reason for the opposing interpretations is an extraneous factor – an overriding consideration involving order of quantifiers in surface structures – that filters out certain latent interpretations provided by the deep structures. In support of this view, it may be pointed out that other sentences that derive from these (e.g., “there are two languages that everyone in the room knows”) may switch interpretations, indicating that these interpretations must have been latent all along. There are other examples that suggest something similar. … (pp. 224-5)

We will return to quantifiers in Section 3; in 1965, linguists had not yet discovered the most serious problems they pose for the Katz-Postal hypothesis.

2.3 The “Standard Theory” and the Garden of Eden Period

During the brief period when Aspects held sway, and its theory had been dubbed by Chomsky the “Standard theory”, there was a rosy optimism that the form of syntactic theory was more or less understood and linguists could concentrate on figuring out the “substantive universals”. Quite a few dissertations were written about the grammar of one language or another, all with deep structures similar to what Chomsky proposed for English in Aspects, and differing only in what transformations applied to make those languages look different from English on the surface. This was also the period when the “Universal Base Hypothesis”, the conjecture that the grammars of all natural languages have the same base rules, was developed independently by McCawley, Lakoff, and Bach; see brief discussions in (Partee et al., 1990, p.556) and (Newmeyer, 1980, pp. 148-50), and more in (Peters and Ritchie, 1973).

In that period, roughly the mid-60’s, before the linguistic wars broke out in full force, generative grammarians generally believed the Katz and Postal hypothesis. The idea that meaning was determined at this “deep” level was undoubtedly part of the appeal of the notion of deep structure beyond linguistics (cf. Leonard Bernstein’s Norton Lectures (Bernstein, 1976)) and probably contributed to the aura surrounding the notion of “language as a window on the mind.”

So around 1965, there was very widespread optimism about the Katz-Postal hypothesis that semantic interpretation is determined by deep structure, and the syntax-semantics interface was believed to be relatively straightforward (even without having any very sophisticated ideas about the nature of semantics.)
3 The “Discovery” of Quantifiers and Expulsion from the Garden

What happened to upset that lovely view? Although of course there were multiple factors, I think it’s fair to focus on one salient issue: linguists discovered quantifiers (Bach, 1968, Karttunen, 1968, Karttunen, 1969, Lakoff, 1968, McCawley, 1971).

It is a surprising historical accident that the behavior of quantifiers with respect to transformational rules familiar from Syntactic Structures and Aspects was not really noticed until the Katz-Postal hypothesis had for most linguists reached the status of a necessary condition on writing rules.

Transformations that preserved meaning (more or less) when applied to names clearly did not when applied to some quantifiers. Clear examples come from “Equi-NP Deletion”, the transformation that applied to (3a) to give (3b).

(3) a. John wants John to win.
   b. John wants to win.

When the identical NPs are names, the transformation is meaning-preserving\(^6\). But if applied to sentences with quantifiers, it would have the unwanted result of deriving (4b) from (4a).

(4) a. Everyone wants everyone to win.
   b. Everyone wants to win.

Similar problems arise for the then-assumed Reflexivization transformation: should (5b) be derived from (5a)? And likewise for the “Conjunction-Reduction” transformation, which would transform (6a) into the non-synonymous (6b).

(5) a. Every candidate voted for every candidate.
   b. Every candidate voted for himself.

(6) a. Every number is even or every number is odd.
   b. Every number is even or odd.

These problems struck at the heart of the Katz-Postal hypothesis and hence at the basic architecture of the Aspects theory. There were two classes of responses by linguists, the Generative Semantics response and the Interpretive Semantics response. The ensuing linguistic wars have been well documented (Harris, 1993, Huck and Goldsmith, 1995, Newmeyer, 1980, Seuren, 1998). Here I want to note how each side could reasonably claim to be following a research program with its roots in Aspects.

The Generative semantics response (Lakoff, Ross, McCawley, Postal, early Dowty, Larry Horn, sometimes Bach) maintained the principle that deep structure should be the input to semantics, and argued (along the lines of Chomsky’s footnote 9 in Chapter 3 of Aspects, but going much farther) that it needs to be deeper, more abstract, more like “logical form” (first-

---

\(^6\) Even with names, later work brought subtle semantic differences to light, e.g. differences between identity de re and identity de se (Chierchia, 1989, Lewis, 1979). We ignore those differences here; the semantic difference between (4a) and (4b) is orders of magnitude greater and was universally appreciated as soon as it was noticed.
order-logic). They were trying to preserve the kind of elegant relation between the deepest level of structure and semantic interpretation that Chomsky had espoused, following Katz and Postal, in *Aspects*. Their adherence to that goal led to many divergences from the syntactic part of the *Aspects* theory; what they held onto and pushed as far as they could was the idea that the underlying level should optimally support semantic interpretation⁷.

The Interpretive semantics response (Chomsky, 1971, Jackendoff, 1972) was to make sure that syntax remained ‘independently motivated’, as Chomsky had insisted in all of his work, and to take examples like those above as evidence that the Katz-Postal hypothesis was mistaken. Chomsky and Jackendoff were ready to give up the principle that an ambiguous sentence should always have two different deep structures, since there did not seem then (nor does there seem now) to be any independent syntactic evidence for syntactic ambiguity for a semantically ambiguous sentence like (7).

(7) Every student answered one question correctly.

The interpretive semanticists preserved more of the syntactic theory of *Aspects*, and gave up the idea of doing all semantic interpretation at the level of deep structure, proposing that different semantic modules may work at different levels, and that quantifier scope and anaphoric relations might be determined at surface structure. The resulting syntax-semantics relationship was architecturally more complex and seemed *ad hoc* to critics.

This note ends here, with the linguistic wars going on in full force, but the story doesn’t.

Although the linguistic wars are now “history”, the issues that were fought over are by no means settled even now. Many very real problems were unearthed in those early post-*Aspects* years. And how best to account for the scope possibilities of quantifiers and other operators has remained one of the most challenging and interesting ones. By the end of the 1980’s there were already at least half a dozen theoretically quite dissimilar proposals for treating the relationship between syntactic structures and the possible semantic scopes of quantifiers in English (see (Partee, 2014) for thumbnail sketches and references), and there are many more now. Argumentation has in the meantime been enriched in many ways -- by a broader range of work on universals and typology, by psycholinguistic studies of syntactic and semantic processing, and by a wealth of studies on the cross-linguistics acquisition of quantificational structures by children.

The *Aspects* model did not survive – no theory does – but the elegance of its architecture of syntax, semantics and phonology sometimes evoke dreams of a lost paradise; a nice dream while it lasted. Actually, fortunately, has been at least as interesting, and the many directions that syntax and semantics have gone in since *Aspects*, and the many discoveries that have been made, have made the syntax-semantics interface one of the most richly flourishing subfields of linguistics.

---

⁷ The generative semanticists generally held to the same questionable principle discussed in footnote 5 above, that semantic relatedness should be reflected in syntactic relatedness at the most abstract or underlying level.
References


**Vocabulary Matters**

**Paul M. Pietroski**  
*University of Maryland*

### 1 Higher Goals

What someone *didn’t* say can be revealing. Consider the following familiar passage.

> Although even descriptive adequacy on a large scale is by no means easy to approach, it is crucial for the productive development of linguistic theory that much higher goals than this be pursued. To facilitate the clear formulation of deeper questions, it is useful to consider the abstract problem of constructing an “acquisition model” for language, that is, a theory of language learning or grammar construction. (pp. 24-25)

As possible contrasts, imagine that Chomsky had instead written (1) or (2).

1. It’s hard to formulate descriptively adequate grammars. But sometimes, it helps to ask how children could acquire the languages they do acquire.
2. Approximating descriptive adequacy is already hard. But in linguistics, the real goal is to explain how children acquire the languages they do acquire.

Alternative (1) would have suggested that the linguist’s job is to describe languages, but that as with many difficult tasks—e.g., achieving wisdom—indirect methods can be useful. And as discussed below, Chomsky’s conception of descriptive adequacy was already quite demanding. Yet he urged linguists to strive for more.

Alternative (2) would have implied that we should aim for a theory of certain biopsychological *processes* that are innately constrained but also input-sensitive. From this perspective, a descriptively adequate grammar for a human language characterizes some state of “knowledge” that any normal child could acquire (given suitable experience), and the linguist’s task is to say how children actually acquire such states given their experience. Some of Chomsky’s remarks are compatible with this construal of his project. But his claims about the

---

1 Page references are to Chomsky (1965) unless otherwise noted. Thanks to Tim Hunter for comments on a previous draft, and to Norbert Hornstein for conversations (over years, and usually over lunch) that made this paper possible.
The competence/performance distinction (pp. 9-10), and the general character of his work, suggest skepticism about the prospects for theories of spatiotemporally located uses of the human language faculty. Episodes of language acquisition are special, in being early uses that constrain later uses of this faculty; but such episodes are still products of many interacting factors.

Indeed, the process of acquiring a language would seem to be more complex than the process of evaluating (3) and (4) for acceptability.

(3) sincerity may frighten the boy
(4) the boy may frighten sincerity

If judging that (4) is less acceptable than (3) involves many disparate factors, this interaction effect may be theoretically intractable (but still useful as data). The process of acquiring a language whose sentences include (3), and perhaps (4), may be far more complicated. In any case, Chomsky speaks of the “abstract problem”—not the real task—of constructing a theory of “language learning or grammar construction.” He says that it is “useful to consider” this problem in order to help formulate “deeper questions,” not that such reflection will lead to a good theory of language learning. But even if language acquisition is an intractable phenomenon, focusing on the abstract problem can be useful. This is because thinking about acquisition highlights the importance, for children and linguists, of the vocabulary used to formulate grammars.

However, children are not little scientists trying to provide theories of languages, given the data and capacities available to adults. Acquisition is better described as selecting a language, from a certain range of options, by using ordinary experience to order the options—or perhaps better still as constructing one or more grammars by using experience to navigate a menu of options. But whatever the details, children presumably encode potential languages in some way that lets experience guide selection as it does. In which case, children are like linguists in an important respect: they employ a vocabulary that supports certain nondemonstrative inferences from “data points” to generalizations that cover unboundedly many cases.

On this overtly mentalistic view, each child can encode any human language, and the coding scheme supports “projections” (cp. Goodman 1954, described below) from parochial courses of human experience to particular languages. In this sense, there is a privileged way of representing the languages that children acquire: the child’s way, which linguists can try to capture. Absent this external standard of correctness for theories of natural languages, which are not governed by explicit stipulations, one might—and Quine (1960) did—worry that any description of such a language is as good (and as bad) as any other, at least within a range of descriptions that meet some minimal conditions of observational adequacy. And while Quine himself assumed a form of behaviorism that Chomsky (1957, 1959) had already criticized, even a 1960s cognitivist might have worried that linguistics was a subject with no objective subject matter.

---

2 Compare n-body problems in physics, many weather patterns, or the production of appropriate novel expressions in communication. Even fifty years later, and allowing for parametric variation in grammatical principles, it seems “impossible to formulate an assumption about initial, innate structure rich enough to account for the fact that grammatical knowledge is attained on the basis of the evidence available to the learner...yet not so rich as to be inconsistent with the known diversity of language” (p. 58).
2 Descriptive Adequacy

In this context, Chomsky suggested a high initial standard of theoretical adequacy.

A linguistic theory must contain a definition of “grammar,” that is, a specification of the class of potential grammars. We may, correspondingly, say that a linguistic theory is descriptively adequate if it makes a descriptively adequate grammar available for each natural language. (p. 24)

A linguist can aim to be like a child whose (internalized) grammar employs a representational format that could be used to encode any language that the child could have acquired. I’ll return to the further desideratum of explanatory adequacy and its relation to the possibility of selecting languages given experience. But thinking about the subject matter of linguistics can lead one to emphasize the acquirability of the languages that linguists study and the importance of representational format, even if one remains agnostic about the prospects for theories of acquisition. Such agnosticism does not render the subject trivial. On the contrary, Chomsky’s notion of a descriptively adequate grammar is already interesting and remarkably demanding.

In particular, he didn’t say that a grammar is descriptively adequate if it specifies all and only the grammatical sentences of the relevant language as such. A proposed grammar for the language is said to be descriptively adequate “to the extent that it correctly describes the intrinsic competence of the idealized native speaker;” where this relative notion, which allows for talk of a grammar being more or less adequate than another, is glossed mentalistically.

The structural descriptions assigned to sentences by the grammar, the distinctions that it makes between well-formed and deviant, and so on, must, for descriptive adequacy, correspond to the linguistic intuition of the native speaker (whether or not he may be immediately aware of this) in a substantial and significant class of crucial cases. (p. 24)

Here, the locution ‘and so on’ is not a casual addition that can be ignored. This passage follows a discussion of how structural descriptions are related to interpretations, as illustrated

---

3 One can get misled by the third paragraph of Syntactic Structures if one ignores the underlined phrase.

From now on, I will consider a language to be a set (finite or infinite) of sentences, each finite in length....The fundamental aim in the linguistic analysis of a language L is to separate the grammatical sequences which are the sentences of L from the ungrammatical sequences which are not sentences of L and to study the structure of the grammatical sequences. The grammar of L will thus be a device that generates all of the grammatical sequences of L and none of the ungrammatical ones. (Chomsky 1957, p. 13, my underlining)

But it helps to distinguish grammars from the aims of linguistic analysis, and to recall the first paragraph.

Syntax is the study of the principles and processes by which sentences are constructed in particular languages. Syntactic investigation of a given language has as its goal the construction of a grammar that can be viewed as a device of some sort for producing the sentences of the language under analysis. More generally, linguists must be concerned with the problem of determining the fundamental underlying properties of successful grammars. The ultimate outcome of these investigations should be a theory of linguistic structure in which the descriptive devices utilized in particular grammars are presented and studied abstractly, with no specific reference to particular languages. One function of this theory is to provide a general method for selecting a grammar for each language, given a corpus of sentences of this language. (Chomsky 1957, p. 11)

Chomsky’s talk of languages as sets, which he elsewhere eschews, was presumably to ease MIT undergraduates into the subject via his discussion of recursion.
with *ambiguous* strings like (5)

(5) I had a book stolen

and the subtle semantic *contrast* between (6) and (7).

(6) I persuaded John to leave
(7) I expected John to leave

By way of highlighting this contrast, note that (8)

(8) I persuaded John that he should leave

is acceptable and a reasonably good paraphrase of (6), while ‘I persuaded that John would leave’ is neither. Yet (9) is acceptable and a reasonably good paraphrase of (7),

(9) I expected that John would leave

while ‘I expected John that he should leave’ is neither. Correlatively (10) is acceptable,

(10) my expectation was that John would leave

but ‘my persuasion was that John would leave’ is not.

It is worth noting explicitly that for Chomsky, a sentence is a string, as opposed to a structured expression. A sentence can have more than one structural description, and thereby have more than one meaning, as (5) illustrates. Given the ubiquity of such structural ambiguity, it is interesting that each of the famous sentences (11) and (12)

(11) John is easy to please
(12) John is eager to please

has *only* one meaning; see Chomsky (1964, p.66). While (11) has the meaning indicated with

---

4 When we “speak of a grammar as generating a sentence with a certain structural description, we mean simply that the grammar assigns this structural description to the sentence” (p. 9); where a structural description “incorporates all information relevant to a single interpretation of a particular sentence” (p. 16). Chomsky (1964) stresses this.

---

The generative grammar of a language should, ideally, contain a central *syntactic component* and two *interpretive components*, a *phonological component* and a *semantic component*. The syntactic component generates strings of minimal syntactically functioning elements (following Bolinger, 1948, let us call them *formatives*) and specifies the categories, functions and structural interrelations of the formatives and systems of formatives. The phonological component converts a string of formatives of specified syntactic structure into a phonetic representation. The semantic component, correspondingly, assigns a semantic interpretation to an abstract structure generated by the syntactic component. Thus each of the two interpretive components maps a syntactically generated structure onto a "concrete" interpretation, in one case phonetic, and in the other, semantic. *The grammar as a whole can thus be regarded, ultimately, as a device for pairing phonetically represented signals with semantic interpretations*, this pairing being mediated through a system of abstract structures generated by the syntactic component. *Thus the syntactic component must provide for each sentence (actually, for each interpretation of each sentence) a semantically interpretable deep structure and a phonetically interpretable surface structure*, and, in the event that these are distinct, a statement of the relation between these two structures. (Chomsky 1964, p. 52, my emphasis)
(11a) as opposed to (11b), (12) has the meaning indicated with (12b) as opposed to (12a).

(11a) It is easy for us to please John
(11b) #It is easy for John to please us
(12a) #John is eager for us to please him
(12b) John is eager that he please us

So a descriptively adequate grammar for English must not only include all six strings among the sentences of English, it must assign structural descriptions which indicate that ‘John’ is the direct object of ‘please’ in (11) and the logical subject of the embedded verb in (12), without also assigning structural descriptions that indicate the unavailable interpretations. Yet such a grammar must still assign two structural descriptions to the ambiguous string (13),

(13) John is ready to please

and three structural descriptions to (14), which has the three indicated interpretations.

(14) Mary saw the boy walking towards the railway station
(14a) Mary saw the boy while walking towards the railway station
(14b) Mary saw the boy who was walking towards the railway station
(14c) Mary saw the boy walk towards the railway station

Sentence (14a) implies that Mary walked towards the station. But (14b) and (14c) do not; they imply that boy walked. Though unlike (14c), (14b) can be used to describe a situation in which Mary saw the boy without seeing him walk; see Chomsky (1964, p.73). Interestingly, (15)

(15) this is the railway station (that) Mary saw the boy walking towards

can only have the interpretation corresponding to (14c)—with the implication that the station is such that Mary saw the boy walk towards it—even though the relative clause in (15) seems relevantly like the ambiguous (14). By contrast, (16) is as ambiguous as (14).

(16) this is the railway station such that Mary saw the boy walking towards it

Yet (17) paraphrases (15), and the interrogative (18) is unambiguous in the way that (15) is.

(17) this is the railway station such that Mary saw the boy walk towards it
(18) what did Mary see the boy walking towards

As Chomsky notes, (18) can be used to ask which thing is such that Mary saw the boy walk towards it; but (18) cannot be used to ask which thing is such that Mary saw the boy who was walking towards it, or such that Mary saw the boy while walking towards it. This suggests a common constraint on the interpretation of relative clauses and questions.

One can invent languages that generate (14-18) in a way that imposes laxer constraints. In

---

5 For discussion of perceptual idioms and the relevance of event variables, see Higginbotham (1983). The constraint on extraction from the relative clause ‘who was walking towards it’ is independently interesting; see Ross (1967).
some of these “English” languages,” both (15) and (18) are ambiguous for competent speakers. In others, (15) is ambiguous, but (18) has only its English meaning—or vice versa. One can also invent languages that impose stricter constraints and assign exactly one meaning to each string, including (14). There are also “alt-English languages” in which (15) and (18) unambiguously have the meaning corresponding to (14a) or (14b). Analogous considerations apply to (5-13).

The general point is that for any given language, a descriptively adequate grammar assigns exactly \( n \) structural descriptions to a string of words—or more precisely, to a string of formatives (p.16)—that can be understood in exactly \( n \) ways, and the grammar assigns typologically distinct descriptions to strings that are understood in typologically distinct ways. These are huge demands that go far beyond the mere requirement of specifying the grammatical sentences, which are just sequences of formatives to which structural descriptions can be assigned.

Indeed, it’s hard to see how the sequences of formatives that happen to be grammatical in (i.e., generated by) English could form an interesting class. Endlessly many of these sequences will be ambiguous; and endlessly many will be like some ungrammatical string in not having a structural description of a certain sort. Prima facie, (19) is grammatical while (20) is not,

(19) was the hiker who lost killed in the fire?
(20) *was the hiker who lost died in the fire?

but only because ‘killed’ can be used in the passive; (19a) is grammatical while (20a) is not.

(19a) the hiker who lost was killed in the fire?
(20a) *the hiker who lost was died in the fire?

The interesting constraint illustrated with (20) is that ‘was’ cannot be understood as associated with ‘lost’, embedded in the relative clause, even though (21) is a fine question; see Ross (1967).

(21) the hiker who was lost died in the fire?

The same constraint applies to (19), whose meaning is very different from that of (22),

(22) the hiker who was lost killed in the fire?

which asks whether the lost hiker killed someone.

The grammaticality of (19) is much less interesting than the fact that (19) cannot have the meaning of (22), just as (20) cannot have the meaning of (21). Having \( n \) structural description is just a special case of having \( n \) (but not \( n+1 \)) structural descriptions; and for these purposes, the difference between zero and one is not theoretically important. The interesting facts concern grammaticality/acceptability on a reading, not the grammaticality/acceptability of strings simpliciter. Given an invented language, like a first-order predicate calculus that can be defined

---

6 See Berwick et. al. (2011) for discussion in the context of recent responses to corresponding “poverty of stimulus” arguments. A descriptively adequate grammar will also assign structural descriptions to many strings that “deviate from well-formedness in certain respects” (p. 3). For example, ‘The child seems sleeping’ is a degraded way of saying that the child seems to be sleeping, not that the child seems sleepy; see Higginbotham (1983).

7 But as a sound, (22) does have a structural description corresponding to (22a), with ‘died’ understood as ‘dyed’. And (22/22a) has this structural description, somewhat degradedly (see note 6), even with ‘died’ understood in terms of death; cp. ‘the hiker was disappeared’. Unsurprisingly, strings of uninterpreted formatives are uninteresting.
as a set of unambiguous well-formed formulae, the notion of a legitimate sequence of symbols may be a good proxy for the notion of a structured expression whose meaning somehow reflects the way in which the structure was generated. But the languages that human children naturally acquire are not remotely analogous in this respect.

Chomsky’s notion of descriptive adequacy reflects these points. To be sure, given any language L, one can talk about the languages that generate the same sentences as L. But if the notion of sentence is not theoretically interesting, then languages that generate the same sentences are not similar (or equivalent) in a theoretically interesting way. We can “say that a grammar weakly generates a set of sentences and that it strongly generates a set of structural descriptions (recall that each structural description uniquely specifies a sentence, but not necessarily conversely).” But the point is not that linguists should try to specify grammars that weakly generate the sentences of English or any other language. On the contrary, as Chomsky notes, the “study of weak generative capacity is of rather marginal linguistic interest” (p. 60).

If a proposed grammar doesn’t even generate the sentences that L generates, the grammar is clearly inadequate as a theory of L. In this sense, one can perhaps show that certain proposals are false without noting the meanings that strings can(not) have. But this epistemic point is no basis for assigning theoretical importance to the weak generative capacity of a grammar/language. Any feature of a system can turn out be evidentially relevant for certain purposes.

3 Deeper Questions

To repeat, providing a descriptively adequate grammar for English would be a significant achievement. Providing a descriptively adequate linguistic theory, which “makes a descriptively adequate grammar available for each natural language (p. 24),” would be a major success. Yet Chomsky says that even strong generative capacity is “not necessarily of any particular theoretical interest.” For at least in principle, grammars can employ different basic vocabularies yet generate the same structural descriptions. Putting this point another way, descriptive adequacy (at least to a significant degree) is a necessary but not sufficient condition for what we really want: a theory that perspicuously represents how humans encode grammars.

When a child acquires English, she does not merely acquire a grammar that agrees with others with regard to the structural descriptions assigned to strings of English words; she does not merely internalize a grammar that strongly generates a certain set. A child’s grammar is formulated in terms of a certain vocabulary that makes it possible to construct a certain range of grammars. So linguists who want a descriptively adequate theory presumably need to aim for the “higher goal” of characterizing the vocabulary that children use to formulate grammars.

On the assumption that this is also the vocabulary in terms of which children evaluate grammars, for purposes of selection, there is little point to aiming for mere descriptive adequacy. And indeed, Chomsky characterizes explanatory adequacy in terms of grammar selection.

To the extent that a linguistic theory succeeds in selecting a descriptively adequate grammar on the basis of primary linguistic data, we can say that it meets the condition of explanatory adequacy….The problem of constructing an explanatorily adequate theory is “essentially the problem of constructing a theory of language acquisition, an account of the specific innate abilities that make this achievement possible”. (25-26, my emphasis)
But there is no assumption here about the feasibility of providing theories of how acquisition unfolds. The added requirement is that descriptively adequate grammars be formulated in a way that makes it possible to choose one, given suitably characterized data.

This is, in effect, an “interface” requirement: the vocabulary in terms of which grammars are formulated must be “connectable” with the vocabulary in terms of which “primary linguistic data” is characterized; cp. Lightfoot (1993). In terms of the initial quote,

Although even descriptive adequacy on a large scale is by no means easy to approach, it is crucial for the productive development of linguistic theory that much higher goals than this be pursued. To facilitate the clear formulation of deeper questions, it is useful to consider the abstract problem of constructing an “acquisition model” for language, that is, a theory of language learning or grammar construction. (pp. 24-25)

the “deeper questions” concern the vocabularies of theorists and children, and the need for a metric in terms of which grammars are evaluable. And while Chomsky formulates explanatory adequacy as a condition on theories that is logically stronger than the demand of descriptive adequacy, there is an important sense in which the same rationale underlies “both” conditions.

The weaker demand would, by itself, be arbitrary; cp. Quine (1960). Why require that a descriptively adequate grammar for any one natural language be formulated in a vocabulary that permits construction of a descriptively adequate grammar for any such language, if not because the “higher goal” is to represent the various languages of this sort in terms of a metric that reflects their underlying commonality in a way that makes it comprehensible how the variation could be correlated with variation in individual experience? Chomsky puts the point compactly by saying that a descriptively adequate theory may “leave unexpressed major features that are defining properties of natural language and that distinguish natural languages from arbitrary symbolic systems” (p. 36). By “adding” the requirement of explanatory adequacy, motivated by reflection on the abstract problem presented by language acquisition, Chomsky connected the more obviously empirical challenges of providing descriptively adequate grammars for particular languages to the more obviously theoretical questions concerning human cognition that had long animated debates between rationalists and empiricists; see Hornstein (2005a, 2005b).

In my view, one towering achievement of Aspects is that it provided a remarkably clear and still unsurpassed conception of the subject matter of linguistics. It made sense of how the method using acceptability judgments could, as it clearly did, lead to rational revision of theories concerning objective natural phenomena. But far from calling on linguists to provide models of complicated spatiotemporally located processes of production, comprehension, or acquisition, Chomsky outlined a project of describing natural languages in the right way—i.e., as kids do.©

4 Gruesome Projections

In stressing the importance of vocabulary and its relation to projection beyond data, Chomsky was influenced by Goodman. So let me connect this discussion to Goodman’s (1954) puzzle.

8 See Fodor (1975) for sustained reflection on this point in the context of other developments. But since linguistic theories are still not explanatorily adequate, it would be rash to conclude that children encode mentalese translations of textbook principles. Likewise, reasonable skepticism about what kids do mentally encode is not a reason for rejecting Chomsky’s conception of the subject matter of linguistics.
Let $t$ be some time in the future, and define the predicates ‘grue’ and ‘bleen’ so that for any entity $e$: $e$ is grue iff $e$ is green and observed before $t$, or blue and not observed before $t$; $e$ is bleen iff $e$ is blue and observed before $t$, or green and not observed before $t$. For any emerald that has already been observed, it is grue iff it is green. So given any observed emeralds that are uniformly positive instances of (23), they are also uniformly positive instances of (24).

(23) all emeralds are green
(24) all emeralds are grue

But while (23) can be confirmed by a (suitably gathered) data set that includes many emeralds all of which are green—and hence grue—such a data set does not confirm (24). So it seems that vocabulary matters for purposes of confirmation.

One might have thought that a generalization of the form ‘all Es are G’ is confirmed by a data set that includes a large sample of Es, gathered in a fair way, if all of the sampled Es are G. Indeed, one might have thought that this sufficient condition for confirmation was nearly trivial, and that challenging it would amount to radical skepticism about “empirical induction.” But as Goodman showed, counterexamples abound, absent restrictions on the kinds of predicates that can be instances of ‘E’ and ‘G’. His own example involves time-indexing: before $t$, observing grue emeralds does not confirm (24), even if the same observations do confirm (23). But reference to times is inessential to the abstract problem that Goodman highlighted.

Let ‘G’ and ‘B’ be contrary predicates: $\forall x(G(x) \supset \sim B(x))$; $\forall x(B(x) \supset \sim G(x))$. Given a third predicate ‘T’, we can define a pair of contrary predicates ‘G*’ and ‘B*’ as in (25) and (26).

(25) $\forall x\{G^*(x) = [G(x) & T(x)] \lor [B(x) & \sim T(x)]\}$
(26) $\forall x\{B^*(x) = [B(x) & T(x)] \lor [G(x) & \sim T(x)]\}$

Alternatively, we can use ‘G*’ and ‘B*’ to define ‘G’ and ‘B’ as in (25a) and (26a).9

(25a) $\forall x\{G(x) = [G^*(x) & T(x)] \lor [B^*(x) & \sim T(x)]\}$
(26a) $\forall x\{B(x) = [B^*(x) & T(x)] \lor [G^*(x) & \sim T(x)]\}$

Suppose the domain consists of things on earth, and ‘T’ applies to things that are north of the equator. Now imagine individuals in the northern hemisphere who endorse generalization (27),

(27) all cows are vegetarians

having noted that every observed cow grazes and never eats meat. If (27) has lots of positive

9 Think of the domain as divided into six regions: (the union of) regions 1, 3, and 5 correspond to (the extension of) ‘T’; regions 2, 4, and 6 correspond to ‘~T’. Let regions 1 and 2 correspond to ‘G’, while 3 and 4 correspond to ‘B’. Regions 5 and 6 correspond to ‘~G & ~B’. Then 1 and 4 correspond to ‘G*’, while 2 and 3 correspond to ‘B*’. The extension of ‘G*’ can be described as a union of intersections: $(\{1, 2\} \cap \{1, 3, 5\}) \cup \{3, 4\} \cap \{2, 4, 6\}$; i.e., $\{1, 4\}$. But likewise for ‘G’: $(\{1, 4\} \cap \{1, 3, 5\}) \cup \{2, 3\} \cap \{2, 4, 6\}$; i.e., $\{1, 2\}$. Similar remarks apply to ‘B*’ and ‘B’. In the context of language acquisition, and the difficulty of learning constraints without “negative data,” it is also worth noting that ‘grue’ and ‘blue’ can be defined so that each of these predicates applies to entity $e$ iff $e$ is blue or $e$ is both green and examined before $t$. Then all green emeralds examined to date are both grue and blue. Or put another way, the extension of ‘grue’/‘blue’ is a superset of the extension of ‘blue’.


instances, and no negative instances, the “inductive leap” seems reasonable—whether or not you know about hemispheres. Since ‘vegetarian’ and ‘carnivore’ are contraries, (27) implies (28).

(28) no cows are carnivores

But if ‘vegetarians*’ and ‘carnivores*’ are defined as in (25) and (26), then generalization (27*)

(27*) all cows are vegetarians*

is unwarranted, despite being like (27) in having each observed cow as a positive instance. For unlike (27), (27*) implies that all cows in the southern hemisphere are carnivores, and hence that (28) is false if there are any cows in the southern hemisphere.

With regard to ‘grue’, one can’t avoid “cherry picking” the data until time $t$. With regard to ‘vegetarian*’, one can’t avoid cherry picking north of the equator. But the vocabulary-relative character of confirmation remains, even bracketing spatiotemporally-indexed predicates. Even if we observe cows from both hemispheres—and thanks to time travel, even cows observed on the day that the last cows died—there will still be many predicates that all the observed cows satisfy. Let ‘I’ mark the division between things that moo in a certain way, and things that don’t. Suppose that all the observed cows happen to moo in this way, which is typical but not exceptionless for cows. A radical skeptic might insist that (27) is unwarranted, given the “biased sample” and the possibility that atypical mooers eat meat. But one can have laxer standards for confirmation without saying that (27*) is equally well-confirmed; where now, vegetarians* are vegetarians that moo in a certain way or carnivores that don’t moo in that way.

Goodman’s “riddle” of induction is to say what distinguishes generalizations that are confirmed by their positive instances—and the “projectible” predicates used to formulate such generalizations—from their gruesome counterparts. So “good” generalizations, which might well occur to us naturally, are often contrasted with “bad” generalizations that we would never think of without help. But one can also view Goodmanian examples as illustrating another point: one person’s reasonable inductive inference is another person’s leap of faith.

Given three Scrabble tiles in a bag, how often would you need to draw a ‘G’ (and put it back) before concluding that each of the tiles was a ‘G’? Reasonable people can differ. But after seeing a ‘G’ drawn ten times, who would conclude that the tiles spell ‘GOD’? The answer is obvious: someone who thinks that the bag must contain an ‘O’ and a ‘D’, even if sampling provides no evidence of this. And such a person could reach the same conclusion after seeing a ‘G’ drawn once. From the perspective of someone whose only background assumption is that the bag contains three Scrabble tiles, inferring ‘GOD’ will seem crazy, not merely rash. For someone who knows that the bag contains an ‘O’ and a ‘D’, drawing again after seeing a ‘G’ is pointless; you might as well check to see if the cow down the street is eating blue emeralds.

Of course, background assumptions can be wrong, and reasonable people are not unduly stubborn. But Chomsky saw that kids just project languages as they do, without regard for how reasonable their projections are from a scientific perspective. That is, kids project gruesomely. Who would conclude, on the basis of listening to adult speech, that (13), (14), and (16) are ambiguous, but that (11), (12), (15), and (18) are not?

(13) John is ready to please
(14) Mary saw the boy walking towards the railway station
(16) this is the railway station such that Mary saw the boy walking towards it

(11) John is easy to please
(12) John is eager to please
(15) this is the railway station (that) Mary saw the boy walking towards
(18) what did Mary see the boy walking towards?

If you’re Chomsky, the answer is obvious: someone who comes to the task with a definition of “grammar”—i.e., a child who has “a specification of the class of potential grammars” (p. 24).

Linguists try to represent this specification by trying to provide an explanatorily adequate linguistic theory. But good scientists do not ignore the possibility of languages that violate the child’s definition. On the contrary, by envisioning “inhuman” languages (e.g., the English+ languages described in section two) and showing that children do not acquire them, linguists can provide scientific justifications for hypotheses about the vocabulary children employ. The child’s vocabulary determines a limited “class of potential grammars” that can be evaluated given ordinary linguistic experience. The linguist’s vocabulary permits description of a much larger class of languages, but also the use of evidence (unavailable to children) to evaluate the scientific hypothesis that children can only acquire the languages that respect certain constraints.

The child’s projections are gruesome, from a scientific perspective, because the child’s “hypothesis space” is constrained by the child’s definition of grammar. This definition supports language acquisition, and in that sense is adequate for a child’s purposes. But the vocabulary of an explanatorily adequate linguistic theory may not be well suited to modeling causal processes acquiring languages. For these scientific purposes, the child’s definition may be inadequate.

We can imagine a mind that deploys the vocabulary of its best physical theory—and the corresponding simplicity metric used to evaluate candidate hypotheses in light of data—in the service of projecting from courses of linguistic experience onto procedures that pair signals of some sort with interpretations of some sort. The class of “languages” this mind can acquire is determined by the vocabulary and simplicity metric it uses to formulate and evaluate physical theories. Whatever the merits of such a thinker, it will not project from courses of experience to grammars in the way that a human child would; and there is nothing scientific about pretending otherwise. One of the great merits of Aspects was that it made this point vividly clear. In my view, this makes Aspects a significant if still under-appreciated contribution to philosophy. The contribution to linguistics is more obvious, though intimately related.

References


LEIBNIZIAN LINGUISTICS

IAN ROBERTS
JEFFREY WATUMULL
University of Cambridge

“Nothing is more important than to see the sources of invention which are, in my opinion, more interesting than the inventions themselves.” (Leibniz, in Pólya 1945: 123)

1 Introduction

The formulation of generative grammar in *Aspects of the Theory of Syntax* (1965) was foundational to the Second Cognitive Revolution—the revival of rationalist philosophy first expounded in the Enlightenment. (That the First Cognitive Revolution of the 17th Century contained “the sources of invention” for generative grammar was appreciated only in retrospect.) Chomsky discussed this historical connection at some length in *Aspects* (see especially Chomsky’s *Cartesian Linguistics* (1966/2009)), identifying Leibniz as one of his precursors in adopting the rationalist doctrine that “innate ideas and principles of various kinds [...] determine the form of the acquired knowledge in what may be a rather restricted and highly organized way” (Chomsky 1965: 48). There are in fact many aspects of Leibniz’s prodigious thought which anticipate core aspects of generative grammar, including some rather specific features of current theory. Below we sketch out what these are, and suggest that we should take Leibniz’s lead in situating the ideas about human nature that generative grammar as a theory of language embodies, and which Chomsky has expounded many times, in a wider philosophical context.

First a very brief introduction to Leibniz’s achievements. It is completely impossible to do justice to the breadth and depth of Leibniz’s work in a short article like this (or even in a long book; Russell 1900/1992). Suffice to say that he was one of the greatest polymaths of all time—a “universalgenie” (a universal genius). Among his achievements are that he discovered calculus (independently of Newton) and other important aspects of mathematics, invented mechanical calculators, propounded a system of rationalist philosophy, formulated his own metaphysics (“the best of all possible worlds”), and made significant innovations in physics, probability theory, biology, medicine, geology, psychology, and what would centuries later become computer
science. He also made a major contribution to the nascent subject of comparative linguistics in that he was one of the first to argue that the Finno-Ugric and Samoyedic families may be connected, thus taking “an early step towards the recognition of Uralic” (Campbell and Poser 2008: 91; this was almost a century before comparative Indo-European linguistics started). His achievements in logic were extraordinary; Kneale and Kneale (1962: 320) say that he “deserves to be ranked among the greatest of all logicians.” In The History of Western Philosophy, Russell observes that Leibniz anticipated much of modern formal logic, as developed in the late 19th and early 20th centuries by Frege and others (including Russell himself). Indeed Russell (1946: 541) states that “Leibniz was a firm believer in the importance of logic, not only in its own sphere, but as the basis of metaphysics. He did work on mathematical logic which would have been enormously important if he had published it; he would, in that case, have been the founder of mathematical logic, which would have become known a century and a half sooner than it did in fact.” (Indeed it was not until the 1950s that mathematical logic enlightened linguistics: “Mathematical logic, in particular recursive function theory and metamathematics, were becoming more generally accessible [in the 1950s], and developments in these areas seemed to provide tools for a more precise study of natural languages as well. All of this I personally found most stimulating” (Chomsky 1955/1975: 39).)

Although this sketch is extremely perfunctory, it conveys something of Leibniz’s staggering achievements. But our aim here is to focus on just a few aspects of this huge legacy, namely: Leibniz’s formal innovations, his metaphysics, and his rationalist epistemology. The first and last of these show direct connections to modern linguistics (as Chomsky pointed out in relation to rationalist epistemology in Aspects, as we have already mentioned):

“Leibniz took Hobbes seriously when [Hobbes] said that reason is nothing but reckoning [computation]. [Leibniz] devoted much of his life to inventing a scheme that would perfect the computations underlying thought, turning arguments into calculations and making fallacies as obvious as errors in arithmetic[...] The idea that intelligence arises from manipulation of symbols by rules is a major doctrine of the school of thought called rationalism, generally associated with Leibniz and Descartes. When the symbols stand for words and the rules arrange them into phrases and sentences, we have grammar, the subject of Cartesian linguistics, which later inspired Humboldt and then Chomsky. When the symbols stand for concepts and the rules string them into chains of inference, we have logic, which became the basis for digital computers, the artificial intelligence systems that run on them, and many models of human cognition” (Pinker 1999: 88-89).

Finally, Leibnizian metaphysics resonates with certain ideas in modern physics, and we will speculate on a connection to cognitive science in a way that we believe, is very much in Leibniz’s spirit.

---

1 It is telling that Leibniz was the subject of the only book the prolific Russell ever wrote about another philosopher.

2 Leibnizian philosophy was an obsession of one the greatest mathematical logicians, Kurt Gödel (see Goldstein 2005). Thus it is unsurprising that when Gödel and Chomsky met, generative grammar was overshadowed by Leibniz’s ghost: “In 1958-9, I spent a year at the IAS. I naturally paid a visit to Gödel. He must have had every book on Leibniz in the Princeton library in his study. As soon as we started talking he instructed me that I was on the wrong track and that I should immerse myself in Leibniz’s work, because that’s where all the answers were about language and meaning” (Noam Chomsky, personal communication).
2 Leibniz’s Formal Innovations

The eminent mathematician and computer scientist Martin Davis (2012) describes the history of computer science as “the road from Leibniz to Turing.” Leibniz devised a calculus ratiocinatar (an abstract computer) operating on a characteristica universalis (a symbolic language). This system is recognizably a universal computational system in the modern (Turing) sense: it is a system which manipulates symbols in a precisely defined (i.e., computable), recursive fashion: “A good century and a half ahead of his time, Leibniz proposed an algebra of logic, an algebra that would specify the rules for manipulating logical concepts in the manner that ordinary algebra specifies the rules for manipulating numbers. He introduced a special new symbol ⊕ to represent the combining of quite arbitrary pluralities of terms. The idea was something like the combining of two collections of things into a single collection containing all of the items in either one” (Davis 2012: 14-15). This operation is in essence formally equivalent to the Merge function in modern syntactic theory (Chomsky 1995); and Merge itself is but a generalization and simplification of the phrase structure rules assumed in Aspects. Leibniz defined some of the properties of ⊕—call it Lerge—thus:

(1) X ⊕ Y is equivalent to Y ⊕ X.
(2) X ⊕ Y = Z signifies that X and Y “compose” or “constitute” Z; this holds for any number of terms.

“Any plurality of terms, as A and B, can be added to compose a single term A ⊕ B.” Restricting the plurality to two, this describes Merge exactly: it is a function that takes two arguments, α and β (e.g., lexical items), and from them constructs the set {α, β} (a phrase). (We can also see that ⊕ shares with Merge an elegant symmetry, as (1) states.) And according to Leibniz’s principle of the Identity of Indiscernibles, if Merge and Lerge are formally indiscernible, they are identical: Merge is Lerge. These functions embody the modern inductive/combinatorial conception of sets: “naturally prior involves the more simple[...]. Prior by nature is a term which consists of terms less derived. A term less derived is equivalent to one [which includes] a smallest number of primitive simple terms” (Leibniz, in Nachtomy 2008: 78). Here the intuition is clearly close to definition by induction, a central aspect of recursive computational systems such as generative grammar (see Watumull, et al. 2014).

In addition to definition by induction, recursive Merge/Lerge realizes the mathematical induction from finite to infinite (e.g., drawing from a finite lexicon, Merge generates an infinite

---

3 “[Leibniz] grasped the concept of having formal, symbolic, representations for a wide range of different kinds of things. And he suspected that there might be universal elements (maybe even just 0 and 1) from which these representations could be built. And he understood that from a formal symbolic representation of knowledge, it should be possible to compute its consequences in mechanical ways—and perhaps create new knowledge by an enumeration of possibilities” (Wolfram 2013: http://blog.stephenwolfram.com/2013/05/dropping-in-on-gottfried-leibniz/). The connections to Chomsky’s ideas are self-evident.

4 Indeed, both Merge and Lerge are purely mechanical (i.e., explicit, generative, deterministic): “all these operations are so easy that there would never be any need to guess or try out anything” (Leibniz), just as a generative grammar “does not rely on the intelligence of the understanding reader” (Chomsky 1965: 4). And the operations Leibniz posited allow us to “proceed to infinity by rule” (Leibniz), just as a generative grammar computes a discrete infinity of syntactic structures; see http://blog.stephenwolfram.com/2013/05/dropping-in-on-gottfried-leibniz/.
set of syntactic objects; see footnote 4). From simple primitives, Nature recursively generates infinite complexity. A complex is generated by iterative “reflection” on simples; “God” reflects on his thinking. “Leibniz [...] presupposes logical simples[...], which he identifies with God’s simple attributes (or God’s simple forms). At the same time, for Leibniz, God is an active mind whose primary activity is thinking and self-reflection. God’s reflections on his simple attributes are mental combinations of his simple forms that produce complex forms. Likewise, God’s reflective operations are iterative, so that he reflects upon his reflections. Thus God thinks the combinations among his simple forms, and more complex concepts arise in his mind. In this view, God combines the simple forms in a natural order—from the simple to the complex—and, in this sense, Leibniz’s system of possibility is both recursive and yields the production of infinite concepts” (Nachtomy 2007: 42-43). Evidently God is equipped with Merge/Lerge. So too are we. Leibniz observes: “The following operation of the mind seems to me to be most wonderful: namely, when I think that I am thinking, and in the middle of my thinking I note that I am thinking about my thinking, and a little later wonder at this tripling of reflection. Next I also notice that I am wondering and in some way I wonder at this wonder” (Leibniz, in Nachtomy 2008: 77). This recursive operation of reflection instances the more general notion of “Reducing reasoning to formal rules”, which “was the underlying basis for Leibniz’s dream of a universal computational language. And it underlay Turing’s achievement in showing that all computation could be carried out on his universal machines” (Davis 2012: 177). And of course, in a decidedly rationalist move, Chomsky has demonstrated that linguistic cognition reduces to computation: “there is no difficulty in principle in programming a computer with [...] a universal grammar of the sort that has been proposed in recent years[...]. I believe these proposals can be properly regarded as a further development of classical rationalist doctrine, as an elaboration of some of its main ideas regarding language and mind” (Chomsky 1968/2006: 73-74). Here we see a profound connection between Leibniz’s thought and modern linguistic theory. But the recursivity of Lerge is but one instance of recursion in Leibnizian philosophy.

Leibniz demonstrated the quadrature of the circle by generating an infinite series from a recursive function: \( \pi/4 = 1 - 1/3 + 1/5 - 1/7 + 1/9 - 1/11 + 1/13 - \ldots \). Leibniz rejected as too weak the Aristotelean (constructivist) notion of potential infinities and accepted the Platonic existence of actual infinities. We can therefore assume that he would not have balked at the formulation of discrete infinity in generative grammar: “A grammar of [a language \( L \)] can be regarded as a function whose range is exactly \( L[.] \) We can consider a grammar of \( L \) to be a function mapping the integers onto \( L \)” (Chomsky 1959: 137, 138). And of course the integers form an actual infinity in Leibniz’s sense—the Platonic sense. More generally, as Chomsky put it in Aspects: “The syntactic component specifies an infinite set of abstract formal objects, each of which incorporates all information relevant to a single interpretation of a particular sentence” (Chomsky 1965: 16). Such unboundedness would have enthralled Leibniz: “I am so in favor of the actual infinite that instead of admitting that Nature abhors it, as is commonly said, I hold that Nature makes frequent use of it everywhere, in order to show more effectively the perfections of its Author” (Leibniz, in Davis 2012: 51). And that Author generates infinite complexity from finite simplicity: “the simplicity of the means counterbalances the richness of the effects” so that in nature “the maximum effect [is] produced by the simplest means” (Leibniz, in Chaitin 2005: 63). This is an informal statement of a fundamental theorem in game theory, finally formalized in the 20th Century.

Indeed Leibniz anticipated important aspects of game theory, particularly von Neumann’s minimax theorem. Game theory is the mathematical theory explicative of goals, logics, strategies,
and related concepts, for which “the fundamental theorem on the existence of good strategies” (von Neumann 1953: 125) is the minimax theorem: “minimize maximum negatives and maximize minimum positives.” This is a good strategy—a principle of optimality— independent of the domain of application because, as Leibniz observed, “the simplicity of the means counterbalances the richness of the effects” so that in nature “the maximum effect is produced by the simplest means.” And “given that things exist,” and given the assumption that “nothing comes about without a sufficient reason[...], we must be able to give a reason why [things] have to exist as they are and not otherwise” (Leibniz, in Chatin 2005: 63; this is the powerful Principle of Sufficient Reason. Observing this principle in the domain of recursive function (computability) theory as applied to generative grammar, the optimal Merge function—and here we are interested in the inductive generation of sets (i.e., the recursive generation of structured expressions)—is minimax (see Watumull and Roberts, Under Review): an effectively unary binary function. Unarity implements minimality (e.g., minimal search, noncounting, etc.): one input suffices to generate an output; but such output is set theoretically trivial (i.e., the potentially infinite subsumption of a single substantive element: α, {α}, {{α}}, ...; {{α}}{{α}}...{{α}}...}—an embedded lexical item). Binarity implements maximality: the capacity to combine two inputs recursively suffices to generate infinitely structured (information carrying) outputs (i.e., {α, β}, {γ, {α, β}}, {δ}, {γ, {α, β}});...; {{δn}, ..., {δi}, {γ, {α, β}}})—a complex sentence or discourse.

The reason to minimize is self-evident (i.e., anything superfluous violates the principle of sufficient reason); the reason to maximize, Leibniz argues, is that it “leave[s] nothing [...] which does not display that beauty of which it is capable” such that a mechanism with the potential to generate infinite complexity ought to realize that potential. Failure to maximize results in a reserve—an unnecessary and hence inexplicable resource—inconsistent with the principle of sufficient reason and perfect design. “When the simplicity of God’s way is spoken of, reference is specially made to the means which he employs, and on the other hand when the variety, richness and abundance are referred to, the ends or effects are had in mind. Thus one ought to be proportioned to the other, just as the cost of a building should balance the beauty and grandeur which is expected” (Leibniz, in Rescher 1991: 196). Nature has “paid” for the unbounded generative capacity of Lerge/Merge, and for Platonized balance thus “expects” and “deserves” unbounded generation. Crudely put, getting “the biggest bang for the buck” is a principle of nature.

Thus in the Leibnizian “best of all possible worlds,” Merge would be propertied with the minimality of unarity and the maximality of binarity. Let us postulate such a minimax Merge as a formally explicit—hence comprehensible and testable—hypothesis. The Minimalist Program (Chomsky 1995) can be seen, in this light, as investigating the hypothesis that the computational system of language is that we would expect to find in a Leibnizian “best of all possible worlds”.

Such a world is a reflection of perfection in that everything ought to be derivable from—explicable in terms of—the simple forms. These are the primitives, the axioms, of nature.

5 “The assertion that nature is governed by strict laws is devoid of all content if we do not add the statement that it is governed by mathematically simple laws[...]. That the notion of law becomes empty when an arbitrary complication is permitted was already pointed out by Leibniz in his Metaphysical Treatise[...]. The astonishing thing is not that there exist natural laws, but that the further the analysis proceeds[...], the finer the elements to which the phenomena are reduced, the simpler—and not the more complicated, as one would originally expect—the fundamental relations become and the more exactly do they describe the actual occurrences” (Weyl 1932: 40, 41, 42). Leibniz and Weyl’s comments apply perfectly to the Minimalist Program: “Some basic properties of language are unusual among biological systems, notably the property of discrete infinity. A working hypothesis in generative grammar has been that languages are based on simple principles that interact to form often intricate structures, and that the language faculty is nonredundant, in that particular phenomena are not ‘overdetermined’ by principles of language. These too
Nothing complex could exist but for irreducible simples. “For this reason, complex thoughts or concepts presuppose constituents. Leibniz indeed presupposes absolutely simple constituents or forms. He writes that ‘nothing can be said of forms on account of their simplicity’ [...]” (Leibniz, in Nachtomy 2008: 76).

The notion that complex thoughts presuppose constituents is axiomatic to Leibniz’s universal computational language. Moreover, the notion has been adduced to argue that not only are natural languages universally constituency-based, but that any language anywhere in the universe would be: we should predict convergence on elementary recursive functions operating on simple constituents; and thus we should predict that “intelligent aliens will be intelligible” (Minsky 1984). However, at the time of Aspects, Chomsky would probably have rejected this hypothesis: “according to the theory of transformational grammar, only certain kinds of formal operations on strings can appear in grammars—operations that, furthermore, have no a priori justification. For example, the permitted operations cannot be shown in any sense to be the most ‘simple’ or ‘elementary’ ones that might be invented” (Chomsky 1965: 55). Leibniz would not have accepted this theory, for in his philosophy nothing is without a priori justification (i.e., everything has a sufficient reason) and such justifications are ultimately based on a metaphysics in which the most complex objects are generated by the most elementary combinatorics (i.e., ⊕) operating over the most simple representations (monads, on which more below).

Interestingly, since Aspects, generative grammar has become increasingly Leibnizian. Already in Lectures on Government and Binding (1981), there were hints that language may be “a system that goes well beyond empirical generalization and that satisfies intellectual or even esthetic standards” (Chomsky 1981: 14). And now, in the Minimalist Program, we can seriously entertain the thesis that the language faculty conforms to “virtual conceptual necessity, so that the computational system [is] in some sense optimal” (Chomsky 1995: 9). If this Leibnizian progression continues, future research will prove language to be governed by absolute conceptual necessity—truly the best of all possible worlds.6

At this point, we should turn to Leibniz’s metaphysics.

3 Metaphysics

It is ironic that the codiscoverer of the calculus should have based his metaphysics not on its continuum—a “labyrinth” to Leibniz’s mind—but rather on discrete substances: monads. For Leibniz, reality is composed of these irreducible immaterial objects—their existence an inescapable logical truth: everything complex is ipso facto composed of simpler things; if those simpler things are themselves extended (i.e., material), then ipso facto they are further reducible; thus it follows that at bottom the simplest of simple substances must be un-extended (i.e., immaterial); “these simple forms are unanalyzable and indefinable [...]. Interestingly, Leibniz identifies these simple forms with the attributes of God. He writes that ‘God is the subject of all

are unexpected features of complex biological systems, more like what one expects to find (for unexplained reasons) in the study of the inorganic world. The approach has, nevertheless, proven to be a successful one, suggesting that the hypotheses are more than just an artifact reflecting a mode of inquiry” (Chomsky 1995: 168).

6 Just as Chomsky’s simplification of phrase structure rules in Aspects revealed greater systematicity, so did Leibniz reveal elegant structure in mathematics by formulating binary arithmetic: “as numbers are reduced to the simplest principles, like 0 and 1, a wonderful order is apparent throughout” (Leibniz); see http://blog.stephenwolfram.com/2013/05/dropping-in-on-gottfried-leibniz/.
absolute simple forms’ [...] and that ‘[a]n attribute of God is any simple form’” (Nachtomy 2008: 76). Substitute “nature” for “God,” for a naturalistic philosophy (i.e., scientific—indeed rational—legitimacy). What are these simple forms? What simply is? What is in need of no explanation? *The truths of mathematics.* Like Leibniz, “Descartes had raised the question of whether God had created the truth of mathematics. His follower Nicolas Malebranche [...] firmly expressed the view that they needed no inception, being as eternal as anything could be” (Dennett 1995: 184). For such reasons, mathematics was foundational to Leibnizian philosophy (and generative grammar). Indeed monads are inherently mathematical (even Platonic).

Monadology implies recursivity: “a theory that explains the status of physical phenomena (together with their relations) in terms of collections of monads mutually representing the objective contents of one another’s representations in a fashion that is infinitely recursive” (McGuire 1985: 215). The simplest forms are binary. Here again we see a link to modern linguistic theory (Merge) and to cognitive science (the computational theory of mind). But Leibniz’s metaphysics, the recursive monadology, led him further: “Leibniz was very proud to note how easy it is to perform calculations with binary numbers, in line with what you might expect if you have reached the logical bedrock of reality. Of course, that observation was also made by computer engineers three centuries later” (Chaitin 2005: 61). Binary enables minimax: “It is the mystic elegance of the binary system that made Leibniz exclaim *Omnibus ex nihil ducendis sufficit unum.* (One suffices to derive all out of nothing.)” (Dantzig 1930/2007: 15). To derive everything from (virtually) nothing allows for a complete explanation.

Laplace commented as follows on how Leibniz linked his binary arithmetic to metaphysics: “Leibniz saw in his binary arithmetic the image of Creation[...]. He imagined that Unity represented God, and Zero the void; that the Supreme Being drew all beings from the void, just as unity and zero express all numbers in his system of numeration” (Laplace, in Dantzig 1930/2007: 15). Here there is a connection with some ideas in contemporary physics: “Leibniz’s vision of creating the world from 0’s and 1’s refuses to go away. In fact, it has begun to inspire some contemporary physicists, who probably have never even heard of Leibniz[...]. What is the Universe made of? A growing number of scientists suspect that information plays a fundamental role in answering this question. Some even go so far as to suggest that information-based concepts may eventually fuse with or replace traditional notions such as particles, fields and forces. The Universe may literally be made of information, they say, an idea neatly encapsulated in the physicist John Wheeler’s slogan: ‘It from Bit’ [Matter from Information]. So perhaps Leibniz was right after all” (Chaitin 2005: 61, 62). If information derives matter, if the mathematics of 0’s and 1’s constitutes reality, if that which requires no explanation—that which is self-explanatory—derives all that does require explanation, then a *theory of everything* is possible. As Leibniz observes: “given that things exist,” and given the assumption that “nothing comes about without a sufficient reason[...], we must be able to give a reason why [things] have to exist as they are and not otherwise.” Thus Leibniz anticipated Chomsky’s (2004) enterprise to go “beyond explanatory adequacy.” Stated generally, to go beyond explanatory adequacy is to give a sufficient reason. Here we speculate, but for generative grammar in particular, the reason

---

7 “[Leibniz’s] typical approach seems to have been to start by trying to create a formal structure to clarify things—with formal notation if possible. And after that his goal was to create some kind of ‘calculus’ from which conclusions could systematically be drawn” (http://blog.stephenwolfram.com/2013/05/dropping-in-on-gottfried-leibniz/). This relates to that idea that “The most interesting contribution a generative grammar can make to the search for universals of language is specify formal systems that have putative universals as consequences as opposed to merely providing a technical vocabulary in terms of which autonomously stipulated universals can be expressed” (Gazdar, et al. 1985: 2).
the language faculty assumes the form it does could be that it too has come from bit: the language faculty is a mathematical object (see Watumull (Under Review) for more on this Leibnizian—indeed Platonist or even Pythagorean—idea). This particular mathematical object (computational system) is obviously encoded neurobiologically, which leads us to rationalism.

4 Rationalism

In Cartesian Linguistics, Chomsky (1966/2009) writes that in the early modern era, rationalism “focus[ed] attention on the innate interpretative principles that are a precondition for experience and knowledge and [emphasized] that these are implicit and may require external stimulation in order to become active[...]. The psychology that develops in this way is a kind of Platonism without preexistence. Leibniz makes this explicit in many places. Thus he holds that ‘nothing can be taught us of which we have not already in our minds the idea,’ and he recalls Plato’s ‘experiment’ with the slave boy in the Meno as proving that ‘the soul virtually knows those things [i.e., truths of geometry, in this case], and needs only to be reminded (animadverted) to recognize the truths. Consequently, it possesses at least the idea upon which these truths depend. We may say even that it already possesses those truths, if we consider them as the relations of the ideas’” (Chomsky 1966/2009: 62-63). This rationalist current flowed into romanticism, as Chomsky observed in Aspects: “Like Leibniz, [Humboldt] reiterates the Platonistic view that, for the individual, learning is largely a matter of Wiedererzeugung, that is, of drawing out what is innate in the mind” (Chomsky 1965: 51). “In the traditional view a condition for [...] innate mechanisms to become activated is that appropriate stimulation must be presented[...]. For Leibniz, what is innate is certain principles (in general unconscious), that ‘enter into our thoughts, of which they form the soul and the connection’. ‘Ideas and truths are for us innate as inclinations, dispositions, habits, or natural potentialities.’ Experience serves to elicit, not to form, these innate structures[...]. It seems to me that the conclusion regarding the nature of language acquisition [as reached in generative grammar] are fully in accord with the doctrine of innate ideas, so understood, and can be regarded as providing a kind of substantiation and further development of this doctrine” (Chomsky 1967: 10).

Chomsky quotes Leibniz at some length in Aspects for the proposition explicitly formulated in generative grammar that we are innately endowed with a competence to generate unbounded knowledge in specific cognitive domains. This knowledge develops if we are presented with the appropriate stimulation. Here Leibniz discusses our related faculty for mathematics: “The truths of numbers are in us, yet nonetheless one learns them [...] by drawing them from their source when we learn them through demonstrative proof (which shows that they are innate),” and thus “all arithmetic and geometry are in us virtually, so that we can find them there if we consider attentively and set in order what we already have in the mind[...]. [In general,] we have an

8 This “Platonism without preexistence” was carried over into modern biology: “Plato says [...] that our ‘necessary ideas’ arise from the preexistence of the soul, are not derivable from experience—read monkeys for preexistence” (Darwin, in Desmond and Moore 1991: 263). Indeed evolutionary psychology—and evolutionary linguistics—does look to the minds of related species for evidence of homologues/analougous to our psychological (e.g., linguistic) capacities.

9 Leibniz is rehearsing Plato’s (Socrates’) idea: “Socrates employs his maieutic methods [...]. Maieutics is the pedagogical method that tries to draw a conclusion out of a mind where it latently resides. As Leibniz remarks of the method, it consists in drawing valid inferences” (Goldstein 2014: 304). It is interesting to consider the similarities and differences when what is being drawn out of the mind is language.
infinite amount of knowledge of which we are not always conscious, not even when we need it[…]. The senses, although necessary for all our actual knowledge, are not sufficient to give it all to us, since the senses never give us anything but examples, i.e., particular or individual truths. Now all the examples which confirm a general truth, whatever their number, do not suffice to establish the universal necessity of the same truth[…]. Necessary truths [...] must have principles whose proof does not depend on examples, nor consequently upon the testimony of the senses”” (Chomsky 1965: 50).

The rationalist theory of language acquisition as running a form of minimum description length algorithm was implicit in Leibniz’s discussion of laws of nature. In connection with data analysis (a central concept in any theory of language acquisition and learnability), Leibniz runs a Gedankenexperiment in which points are randomly distributed on a sheet of a paper and observes that for any such random distribution it would be possible to draw some “geometrical line whose concept shall be uniform and constant, that is, in accordance with a certain formula, and which line at the same time shall pass through all of those points” (Leibniz, in Chaitin 2005: 63), as with a Lagrangian interpolation. For any set of data, a general law can be constructed. “How can we decide if the universe is capricious or if science actually works? And here is Leibniz’s answer: If the law has to be extremely complicated (‘fort composée’) then the points are placed at random, they’re ‘irrégulier,’ not in accordance with a scientific law. But if the law is simple, then it’s a genuine law of nature” (Chaitin 2005: 63). In Leibniz’s words, the true theory is “the one which at the same time [is] the simplest in hypotheses and the richest in phenomena, as might be the case with a geometric line, whose construction was easy, but whose properties and effects were extremely remarkable and of great significance” (Leibniz, in Chaitin 2005: 63). This Leibnizian logic is the logic of program size complexity or, equivalently, minimum description length, as determined by some evaluation/simplicity metric (see Chomsky and Halle 1968: Chapter 9).

“Leibniz had all the pieces, he had only to put them together. For he worshipped 0 and 1, and appreciated the importance of calculating machines. [Leibniz could have anticipated Chaitin by understanding] a scientific theory as a binary computer program for calculating the observations, written down in binary. [There exists] a law of nature if there is compression, if the experimental data is compressed into a computer program that has a smaller number of bits than are in the data that it explains. The greater the degree of compression, the better the law” (Chaitin 2005: 64). This logic applies straightforwardly to language acquisition.

Language acquisition can be represented as navigating a hierarchy of parameters computing a series of decision points (see Roberts 2012, Biberauer and Roberts 2013 and references given there). In these hierarchies, Leibniz’s notion of “the simplest in hypotheses and the richest in phenomena” is represented in the minimax principle of feature-economy (minimize formal features) combined with input generalization (maximize available features). The parameterized language—the steady state—is attained by traversing a path in the hierarchies: the parameterized language is by definition the shortest possible (if described in bits), consistent with Leibnizian program size complexity. As Berwick (1982: 6, 7, 8) puts it, a “general, formal model for the complexity analysis of competing acquisition [...] demands [can be] based on the notion of program size complexity[—]the amount of information required to ‘fix’ a grammar on the basis of external evidence is identified with the size of the shortest program needed to ‘write down’ a grammar. [Formally, this] model identifies a notational system with some partial recursive function Φi (a Turing machine program) and a rule system as a program p for generating an observed surface set of [primary linguistic data]. Like any computer program, a program for a rule system will have a definite control flow, corresponding roughly to an augmented flowchart
that describes the implicational structure of the program. The flow diagram specifies [...] a series of ‘decision points’ that actually carry out the job of building the rule system to output. [The] implicational structure in a developmental model corresponds rather directly to the existence of implicational clusters in the theory of grammar, regularities that admit short descriptions. [T]his same property holds more generally, in that all linguistic generalizations can be interpreted as implying specific developmental ‘programs’”. Thus the complexities of language acquisition and linguistic generalizations are determined by simple programs—a Leibnizian conclusion evident in the theory of parameter hierarchies. More generally, “Leibniz, like Hobbes (who had influenced him), was ahead of his time in recognizing that intelligence is a form of information processing that needs complex machinery to carry it out. As we now know, computers don’t understand speech or recognize text as they roll off the assembly line; someone has to install the right software first” (Pinker 2002: 35).

In the rationalist view of language acquisition, a generative grammar is genetically preprogrammed and activated by primary linguistic data. In Aspects, Chomsky “follow[s] Leibniz’s enlightening analogy, [with] ‘the comparison of a block of marble which has veins, rather than a block of marble wholly even, or of blank tablets, i.e., of what is called among philosophers a tabula rasa. For if the soul resembled these blank tablets, truths would be in us as the figure of Hercules is in the marble, when the marble is wholly indifferent to the reception of this figure or some other. But if there were veins in the block which should indicate the figure of Hercules rather than other figures, this block would be more determined thereto, and Hercules would be in it as in some sense innate, although it would be needful to labor to discover these veins, to clear them by polishing, and by cutting away what prevents them from appearing. Thus it is that ideas and truths are for us innate, as inclinations, dispositions, habits, or natural potentialities, and not as actions; although these potentialities are always accompanied by some actions, often insensible, which correspond to them’” (Chomsky 1965: 52). So, like a figure revealed in marble (the Leibnizian analogy), or a geometrical theorem elicited from an untutored mind (the Socratic example), the form of linguistic knowledge—a generative grammar—is predetermined, but its expression is contingent. (Hence did Leibniz distinguish necessary necessity from contingent necessity.)

5 Conclusion

Contra Leibniz, it is conceivable that some things exist for no reason—things for which there is no explanation—but to conclude so is in general logically impossible (modulo some cases in mathematics): it cannot be proved that an “unsuccessful” search of the infinite space of possible explanations has been exhaustive; it cannot be proved that our search will never succeed (i.e., converge on an explanation); this is analogous to the incomputability of the halting problem and Kolmogorov- Chaitin complexity. But we should try, as in Hilbert’s dictum: Wir müssen wissen. Wir werden wissen. This, it seems to us, is an apt description of the program for linguistic theory set out in full for the first time in Aspects, a program which has set linguistics firmly in a venerable rationalist position, both epistemologically and metaphysically. We seek theories that are not merely observationally and descriptively adequate. Nor are we satisfied with explanatory adequacy. The desideratum, implicit in Aspects and explicit in the Minimalist Program, is to go beyond explanatory adequacy and give a sufficient reason—a reason that would satisfy Leibniz—for why language assumes its unique form—perhaps the best of all possible forms.
References


Watumull, J. Under Review. Language and the Computable Universe Hypothesis.


Watumull, J. and I. Roberts. Under Review. Syntax all the way up.

1 Preamble

My first introduction to syntax was in the fall of 1962, in Zellig Harris’s course at Penn. Harris, a quiet man, was nondescriptly dressed, with eyes that seemed blurred, from looking over vast horizons. There was very little to read on transformational grammar – I doubt that there were more than a few dozen researchers in the world who thought in terms of kernel sentences and ways of transforming them. I have no reason for a feeling that I had in this class, of around twenty of us, in a dusty room in a ramshackle house on Walnut Street – the feeling that we were on the burning edge of syntactic understanding. None of us talked very much – we just listened to Harris talk about syntax from the bottomless understanding of language that the decades of work had given him. I was certainly in awe of him, and I think anyone who felt differently would have been out of it.

Maybe halfway through the course, after I had, by dint of ceaseless reading and thinking, studying incessantly, the first time I had ever worked seriously on linguistics — I had gotten a taste of how beautiful was the system that he could see, and was helping us to see along with him.

And one day in class, he said – I believe – in these very words, “Some transforms of a sentence are more noun-like than others.”

I don’t remember that he gave us any reason for this statement, and I don’t know if there was anyone else in our class, into whose heart this short assertion had also gone all the way through them – all I know is that I have never forgotten this sentence. And in the complement system of English, which I would not be able to get a good grasp of for perhaps ten years, I began to find nets of phenomena which would serve as the evidence that Harris had never given. I gathered them all in a monograph to be published in 1974, under the title “Nouniness.”
I was only to stay at Penn for a year and a half; I finished my Masters degree there and then went on to the turbulent waters of MIT, in January 1964. Jerry Katz and Paul Postal’s groundbreaking book was available in Xerox form, all of us students were abuzz with the rumors of the new book that Chomsky was writing – the book that this volume is celebrating. We all of us could see very clearly a new way of interconnecting syntax, phonology and semantics. Deep structure! It is hard to remember what a huge breakthrough in thinking Aspects represented. We felt that we had a powerful new tool that would revolutionize thinking about syntax. And that it did.

But in this vision there was no smidgin of room for Harris’s insight. I joined in the celebration, I became a Chomsky lieutenant, we all worked like beavers, it was a fertile time.

And after Peter Rosenbaum’s thesis was finished and eagerly put into use in the mid-sixties there were all kind of heady loose ends to follow up on. Working out the details of Peter’s thesis led me back to noun-likeness – I wrote a number of papers on what I called non-discrete grammar, and I began to view grammaticality itself as a scalar entity, one that could definitely be subtracted from, and perhaps even added to, and if grammaticality can be lessed and mored, could even noun-likeness be? Not in Aspects, and in the many papers which it gave rise to, but the seed that Zellig Harris’s throwaway line had planted in me would not be denied. Hence the following lucubrations on some of the problems that such a squishy way of thinking inevitably brings with it.

2 Squishiness

In an antediluvian squib (Ross (1970)), I called attention to the fact that there is a difference between two types of idioms with nominal(oid) objects: some objects can trigger pronominalization, some can’t (cf. (1)).

(1) a. I stubbed my toe, so the doctor will have to look at it.
   b. *You can take your time, if you like, but I doubt that you'll value it.

While I then thought of things in an all-or-none way, I would now be inclined to believe that intermediate cases could be found, in line with my general contention that things in syntax (and elsewhere in linguistics, for that matter) are rarely black/white, but are rather usually of varying shades of grey.

Thus in the present case, I would argue that in the idiom to hold one's breath, the object NP(?) is less pluralizable than toe in (la), but more so than time in (lb). For me, pronominalization is possible for breath with a shared subject, but not with a different one. Thus (2a) works, but not *(2b).

(2) a. Bellwether held his breath, and then (he) released it.
   b. *Bellwether held his breath, so that I could measure its exhalatory velocity
      with the miniaturized anemometer I grafted onto his pharynx.

Thus I envisage an implicational hierarchy along the lines of (3).
to stub one's toe

to hold one's breath

to lose one's way

to take one's time

to stub one's toe

In passing, I note that this hierarchy may play a role in accounting for why only some of these nouns can be modified by the passive participles of their associated verbs:

(4) a. A stubbed toe can be very painful.
     b. *Held breath is usually dank and fetid when released.
     c. ** A lost way has been the cause of many a missed appointment.
     d. ***Taken time might tend to irritate your boss.

Yet another way in which this hierarchy seems to display itself is in interaction with Gapping, the optional rule which can elide a repeated verb in a coordinate clause:

(5) a. I stubbed my toe, and she stubbed hers.
     b. I held my breath, and she held hers.
     c. ?I lost my way, and she lost hers.
     d. *I took my time, and she took hers.

And yet one more: the object nouns (?) in (3) are less and less incorporable, as we see in (6).

(6) a. Please make the children all wear steel-tipped boots, to minimize the danger of toe-stubbing.
     b. ?Prolonged breath-holding may lead to an exalted state.
     c. *Way-losing in the woods leads to merit badge cancellation.
     d. *Undue time-taking at the water-cooler will surely rile Mr. Grarlsh.

Mutatis mutandis, the same is true for Object raising (a.k.a. Tough-Movement):

(7) a. i. To stub one’s toes in the dark is easy ➔ via Object raising
     ii. One’s toes are easy to stub in the dark.
     b. ?One’s breath is harder to hold under the water than above it.
     c. *One's way is almost impossible to lose in a GPS-equipped new Solara™.
     d. *One’s time is impossible to take on the freeway.

However, the main point of this squib is the interaction of the hierarchy in (3) with a rule which I will call Pluralization. With a plural subject, the higher up an idiom is on the list in (3), the less grammatical is the sentence with a singular NP in object position. Compare the sentences in (8):
(8)  a. Jeanne and Minerva stubbed their [toes / *toe].
    b. Jeanne and Minerva held their [breaths / breath].
    c. Jeanne and Minerva lost their [**ways / way].
    d. Jeanne and Minerva took their [**times / time].

It does not suffice merely to say that for stub one’s toe, PLURALIZATION is obligatory, that it is blocked for take one’s time and lose one’s way, and optional for hold one’s breath. For in this last case, there seem to be different conceptualizations associated with the presence or absence of the plural on breath. With singular breath, it is required (for me) that the speaker perceive that there is one event which causes Jeanne and Minerva to either hold their breath physically, i.e., to stop breathing for a while, or, in the metaphorical sense, merely to wait in suspense, and very intently. In short, Jeanne and Minerva are conceptualized as doing this breath-stopping waiting, together, at least at the same time, if not at the same place.

By contrast, while breaths admits of this joint reading (as I hear the sentence), this plural also allows for a reading in which there are two, non-simultaneous, waitings. As would be the case in a context like (9).

(9)  As the swimming teacher went slowly down the list of names, checking each child's breath-holding ability individually, it turned out that Jeanne and Minerva had held their [breaths / ??breath] the longest of any of us.

Another way of forcing a non-joint reading, as noted in Lakoff and Peters (1969), in which the semantactic consequences of jointness are gone into in depth, is to append the quantifier both:

(10)  Both Jeanne and Minerva held their [breaths / *breath].

For me, however, the clearest ungrammaticality results when one collapses such sentences as (11a) into (11b) by means of the operation that produces/sanctions the adverb respectively:

(11)  a. Jeanne held her breath on Tuesday, and Minerva held her breath on Wednesday.
    b. Jeanne and Minerva held their [breaths / *breath] on T

I have been discussing these contrasts in jointness as if they were to be accounted for by a semantactic rule which, under conditions of non-joint interpretation, changes a non-plural form breath (as in (11a)) into a plural one (as in (11b)). Such a view of matters is by no way forced by the facts – it merely is my personal lineage as a transformationalist speaking. Those whose predilections and training have inclined them to a descriptive apparatus involving filters or interpretive rules should encounter no difficulties in constructing a functional equivalent for PLURALIZATION. Nor are they intended to, for the main purpose of this note is not to champion some theoretical framework, but rather to call the attention of all researchers to a shared problem.

I will close with some final observations about the object nouns in (3). In related uses of these nouns, even when they are not in construction with the idiomatically collocated verbs of (3), these nouns seem to differ with respect to how well they can occur in the plural. This becomes clear from such data as those I cite in (12).
(12) a. Hans stubbed both big toes. But since these toes are crucial equipment for a rock-climbing instructor, he'll probably be fully compensated by his insurance company.
   b. Even though pearl-divers in the laboratory were able to hold their breath 3 times longer than normal citizens, pneumometer tests performed on the breaths from both groups of subjects revealed no significant differences in fetidity.
   c. The way to Pike's Peak through the city is far less scenic than the way through the stockyards, but the two ways are about equal in number of beer cans per square meter.
   d. Even if you take your time slicing the carrots and I take my time peeling the onions, these times will still feel like work.

That is, it would be refreshing if the declining acceptabilities in (12) could provide a basis for the differences which have formed the main topic of this paper, and I am hopeful that such a demonstration will one day be feasible.

There is an extremely important issue lurking in the wings here – the question of the conceptual, or possibly perceptual, basis for the count/mass distinction. Let us return to the contrast between (8a) and (8b), which I repeat for convenience.

(8) a. Jeanne and Minerva stubbed their [toes / *toe].
   b. Jeanne and Minerva held their [breaths / breath].

The question which this contrast raises, in my mind, is why we refuse to perceive a simultaneous toe-stubbing (say, one in which Jeanne and Minerva both kick a rock at the same time, as part of a prearranged plan) as codable with a singular toe, as in (11).

(13) On Tuesday, June 9, at 5:17 a.m. precisely, Jeanne and Minerva deliberately stubbed their toe together, precipitating a constitutional crisis unparalleled in recent legal history.

It seems unsatisfying to me to rejoin along the lines of (14):

(14) “But there are two physically distinct toes (except in the rather grotesque case of Siamese twins)! Obviously, it was not one physical object that was injured, but two. Hence the plural.”

The reason is that the way I conceive of the referent of her breath in (15):

(15) Jeanne held her breath.

is as a physical, bounded entity: the gaseous contents of Jeanne’s lungs. To be sure, this is not a physical object, but rather, as we “know” from physics, a collection of molecules. Still, English does not scruple at viewing it as a singular entity, as we can see (in (2a)). And obviously, the set of molecules in Jeanne’s lungs is not the same as the set in Minerva’s, so why should we be able to “fuse” the two distinct volumes, as it were, in the case of a simultaneous breath-hold, to say (16)?

(16) Jeanne and Minerva held their breath together.
3 Prototypical Objects

I do not think that we can look to physics for an answer here. Rather, I believe that what is at issue is a psychological matter: what are the properties of prototypical objects? The provisional answer I would suggest appears in (17).

(17) Objects are prototypically (for a discussion of this crucial psychological, and linguistic, notion, cf. Rosch (1975) and Lakoff (1987)).

a. Solid
b. Spatiotemporally contiguous
c. Not aggregates (like piles, sets, etc.)
d. Not inalienable parts of larger objects (thus elbows and necks are not great objects, though toes and fingers and perhaps tongues are, perhaps because they protrude, seem to be far closer to attaining object-citizenship).

We are willing to depart from all of these: we refer to some spatially-bounded liquids as objects (teardrop, lake, river), and volumes of gas (cloud, column of smoke), and sometimes temporally but not spatially contiguous event parts, as in a joint breath, and even, wildly, in cases of morphological regeneration, as Postal (1976) has pointed out – cf. (18), in which the it refers to an object (?) that is neither spatially nor temporally contiguous with the chameleon’s tail.

(18) My pet chameleon got his tail, caught in the door, but it, will grow back.

But none of these are garden-variety, prototypical objects, and when we depart from the prototype, we find that certain morphosyntactic properties which go with the psychological prototype, such as those in (19), may be absent.

(19) The Count Noun Syndrome (in part)

Count nouns can

a. be pluralized
b. be modified by numbers and many/few, and each
c. trigger pronouns
d. not occur articlelessly in the singular (I kissed *(a) dog.) etc.

What appears to be beating the drum which the constructions that I have been discussing dance to is a gradual departure from the experientially-based prototype in (19) – thus a toe is a better match to the prototypical object than is a breath, and a breath (which is still physical, anyway) is better than is a way – whose physicality or not I will leave to my colleagues in philosophy to debate) and way (which is visualizable, anyway) is better than time. So far, so good, perhaps.

What I have yet to understand is how the factors in (19) are arranged – why does the more or less monotonic decline in experienceability of the nouns in (3) pick one or the other of the
morphosyntactic properties in (19), and the others that the discussions of this paper presuppose, as criterial? Tersely, why all the fuss about pluralizability?

To sum up the discussion above, I think that the following squish\(^1\) is adequate, to the limited extent that I have been able to explore the phenomenon to date.

(20) | Less Restrictive |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ven N</td>
<td>Gapping w/ idiom</td>
<td>Plural</td>
<td>Pronominalizability</td>
</tr>
<tr>
<td>a. stub one's toe</td>
<td>OK</td>
<td>OK</td>
<td>OBL</td>
<td>OK (cf. (1a))</td>
</tr>
<tr>
<td>b. hold one's breath</td>
<td>*</td>
<td>?</td>
<td>“OPT” (but cf. (8)-(11))</td>
<td>OK w/= subject, (cf. (2))</td>
</tr>
<tr>
<td>c. lose one's way</td>
<td>*</td>
<td>?*</td>
<td>BLOCKS</td>
<td>%(^2)</td>
</tr>
<tr>
<td>d. take one's time</td>
<td>*</td>
<td>?*</td>
<td>BLOCKS</td>
<td>*</td>
</tr>
</tbody>
</table>

(20) | OK | OK | OBL | OK (cf. (1a)) | OK |
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>b. hold one's breath</td>
<td>*</td>
<td>?</td>
<td>“OPT” (but cf. (8)-(11))</td>
<td>OK w/= subject, (cf. (2))</td>
<td>OK</td>
</tr>
<tr>
<td>c. lose one's way</td>
<td>*</td>
<td>?*</td>
<td>BLOCKS</td>
<td>%(^2)</td>
<td>??</td>
</tr>
<tr>
<td>d. take one's time</td>
<td>*</td>
<td>?*</td>
<td>BLOCKS</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

(cf. (4)) (cf. (5)) (cf. (6)) (cf. (1b)) (cf. (12))

The problem for future research, as I see it, it to provide an explanation for the ordering of the columns of (20), assuming, that is, that the basis I have suggested for the explanation of the rows – namely, departure from the prototypical notion of physical object – can be made to stand up under detailed scrutiny.

4 How Nouns Lose It

And there is a more general problem, which I can only indicate here: how do nouns lose their identity? What I am thinking of can be suggested by the facts in (20) and (21).

(21) a. Tom bought a set of burglar's tools.
    two sets of burglar's tools.
    *a set
    ?a setta

b. Tom bought a number of burglar's tools.
    * two numbers of burglar's tools.
    *a number
    *a numbera
c. Tom bought a couple of burglar's tools.
   * two couples of
   a couple
   a coupla

d. i. Tom bought a bunch of burglar's tools.
   * two bunches of
   * a bunch
   a buncha

ii. There is a bunch of pears in the fridge.
   * wine
   a buncha {pears/wine}

e. i. Tom bought a lot of burglar's tools.
   * two lots of
   * a lot
   a lotta

ii. There is a lot of []? pears ≤ / OK wine] in the fridge.
   [OK pears/ OKwine]

(22) a. It is on the top of the box.
   on top of the box.
   * ontop the box
   (but cf. atop)

b. It is * in the front of the box (* w/ the sense of “before the box”)
   in front of the box
   * [infront / affront] the box

c. It is * in the side of the box (* w/ the meaning within the box)
   * in side of the box
   inside of the box
   inside the box

d. It is * by the side of the box (* w/ the desired meaning of next to)
   * by side of the box
   * byside of the box
   * beside of the box
   beside the box

It should be pretty clear, intuitively, what is going on in these two examples. In (21), we see a number of collective nouns which are in various: stages of ego-loss. In (21a), it is clear that we simply have two full nouns, set, and (burglar's) tools, while in (21e), the original noun lot, which
Nominal Decay

originally denoted a two-dimensional array, as in a lot of stamps (cf. Pynchon’s The Crying of Lot 49), has vanished entirely, as we can see from the fact that it now accepts mass nouns as objects (a lot of wine), which would be deviant if lot still had only its “array” meaning. Bunch is, for some speakers, on the road down the tubes: when it has contracted with its following of, it too can modify (?) mass nouns (cf. (21di)). And it is already so far gone that it can no longer be pluralized: cf. (21di). Of all the nouns in (21), only the nouniest, set, retains this property (but why is lots of OK?). Although I have not indicated this property in (21), the higher the noun is located in (21), the more easily (and variegatedly) it can be modified adjectivally: [a large number of / *a large lot] of pears.

In (22), we find a similar slide toward ego-loss for such spatial nouns (N_space) as top, front, and side. Without going into details, it is obvious that the stages in this nominal decay are those sketched in (23).

(23) A Scenario for N_space Doom

a. Loss of article preceding N_space: first optional (cf. top), then obligatory (cf. all other examples in (22))

b. Fusion of N_space with preceding locative preposition (obligatory for side)

c. Loss of following of (optional for inside, obligatory for beside)

Some N_space have slid so far that we only find them as bound forms: behind, beneath and underneath; and between. Although I have not chronicled this factor in (22), it is my belief that the “application” of (23a) – the loss of the definite article – is correlated with unpluralizability: cf. on the tops of the boxes vs. *on tops of the boxes.

It would be tempting to conclude that pluralizability correlates with some semantico-pragmatic notion like “usable to refer with.” However, there are counterexamples: plural nouns which seem not to refer to plural sets: lots of wine, or she is keeping tabs on him. Thus I think that a lot of careful work will be necessary here, to remove the chaff from the undeniable grain of truth in such an idea.

So to return, for a brief farewell, to the problem raised by the dwindling compositionality of the idioms in (3), and to the mystery surrounding the ordering of the columns in the squish of (20), it seems we are in deep waters here. We must, as I have argued in a buncha papers on squishes, have a flexible set of metalinguistic predicates, so that we will be able to talk about the mores and lesses of idiomaticity, and of the egohood of nouns. Whether or not we will unearth a single generalized “scenario” for ego-loss in nouns in a wide range of categories remains an enticing, but open, beckon. Schön wär’s.

Meet you over there.

References


FEATURES FROM ASPECTS VIA THE MINIMALIST PROGRAM TO COMBINATORY CATEGORIAL GRAMMAR

NEIL SMITH
ANNABEL CORMACK
University College London

1 Background

One major contribution of Aspects (Chomsky 1965) was to initiate the development of a theory of syntactic features. There is no use of features, either syntactic or morphophonemic, in Chomsky’s earliest work The Morphophonemics of Modern Hebrew (1951/79); they do not appear in his monumental The Logical Structure of Linguistic Theory (1955/75); nor in Syntactic Structures (1957) (except for category labels); nor, except for phonological distinctive features, in Current Issues in Linguistic Theory (1964). In Aspects features play a crucial role. Chomsky suggests (p.214) that “We might, then, take a lexical entry to be simply a set of features, some syntactic, some phonological, some semantic”. But even here the use of features was somewhat haphazard, apparently unconstrained, and a terminological mess. In this contribution we want to trace – in broad outline – the development of a theory of morphosyntactic features and propose a fully Merge-based and parsimonious version of such a theory. Our work is intended as a contribution to Chomsky’s Minimalist Program, pursuing the same ends on the basis of some different background assumptions. These assumptions and the specific analyses derived from them have been developed using an austere version of Combinatorial Categorial Grammar, but are of wider applicability.

All current theories use phonological features and ‘semantic’ features which we will largely ignore except to point out that the term ‘semantic’ is used in radically different ways to cover truth-theoretic, conceptual or encyclopaedic distinctions. We will rather concentrate on Syntactic, Morphological and Morphosyntactic features, each of which, depending on the definition of the last category, also occurs in all theories. Our pre-theoretic intuition is that the syntactic features of a lexical item are things like category V, and the encoding of transitivity, which have no immediate morphological effects; morphological features are things like [3rd

1 Hereafter references to Chomsky’s work will be cited only by date.
Declension] which have morphological but no syntactic repercussions (the last being referred to by Chomsky as a morphosyntactic feature); and morphosyntactic features are things like $[\pm\text{Common}, \pm\text{Count}]$ which have morphological effects with syntactic implications e.g. for agreement.

2 Aspects and On

In the 1975 Introduction to *The Logical Structure of Linguistic Theory* Chomsky says (1955/75:17) “Another modification in the ATS (Aspects) theory was the development of a system of syntactic features … permitting a sharp separation of the lexicon from the remainder of the phrase-structure grammar”. By reconceptualizing the nature of lexical insertion he was restricting the ‘categorial component’ to a context-free grammar and making the scope of the new features much wider. There were now Syntactic and Morphological features and their interaction in a form which presaged the later development of Morphosyntactic features. One implication of this relation can be seen in the remark: (1965:86-87) “many of the grammatical properties of formatives can now be specified directly in the lexicon, by association of syntactic features with lexical formatives … In particular, morphological properties of various kinds can be treated in this way”. He further observes (1965:171) that “we can restate the paradigmatic description directly in terms of syntactic features”. That is, a lexical item such as the German *Brüder* can be associated directly with the set of features (*masculine, genitive, plural* in his example) which characterize a cell in a nominal paradigm. To capture the relation to syntax, rules were simplified by replacing $[+\text{TRANS}]$ with $[+\text{___ NP}]$, where the latter is a new construct, a ‘contextual feature’, subcategorizing the verb in terms of the linear context it can occur in.\(^2\) Such features entered into Phrase structure rules, and appropriate combinations were bundled together as ‘syntactic’ features on a lexical head (1965:107), with the two being connected by Conventions for Complex Symbols (1965:102).

 syntactic features were subdivided into contextual features, and ‘inherent’ or ‘intrinsic’ features, (such as $[+\text{Common}]$ or $[-\text{Count}]$) whose cross-classificatory property motivated the development of ‘complex symbols’ (a set of specified features) such as the underlined part of: $N \rightarrow [+N, \pm\text{Common}]$). Contextual features were in turn divided into Subcategorization features such as $[+\text{NP}]$ or $[+\text{Det}]$, and Selection features, determining the semantically felicitous environments in which items could occur. For example, to account for the acceptability of a sentence like “Sincerity may frighten the boy” and the anomalous nature of “The boy may frighten sincerity”, the verb *frighten* was assigned appropriate features to require an [abstract] subject and an [animate] object. There was subsequently prolonged discussion as to whether the ‘syntactic’ features involved were really syntactic or rather semantic. The consensus gradually emerged that they were semantic, to be replaced later still by the view that because such constraints could be over-ridden in pragmatic exploitation of language, they were not part of the grammar at all.

The next major development in feature theory came with “Remarks on Nominalization” (1970). In this paper Chomsky postulated (1970:215) “a feature $[+\text{cause}]$ which can be assigned

\(^2\) At this period, Chomsky talks of “strict subcategorisation of Verbs in terms of the set of syntactic frames in which $V$ appears”, noting that “Verbs are not strictly subcategorized in terms of Subject NPs or type of Auxiliary, apparently.” Current usage rather speaks of a Verb as being subcategorized for an NP object.
to certain verbs as a lexical property”, and made the radical suggestion (1970:208) that “We might ... eliminate the distinction of feature and category, and regard all symbols of the grammar as sets of features” ... leading to the possibility of “complex symbols of the form [+def, +NP]”, to describe a phrase. Concomitantly, features could now be associated with non-lexical categories such as an article. X-bar theory for lexical heads was formulated (1970:210), introducing the notion of ‘specifier’ of an X’ phrase. The feature system was used in determining well-formed Deep Structures, and to specify the correct input for Transformations.

Neither Lectures on Government and Binding (1981) nor Barriers (1986) developed syntactic feature systems in any depth, and the next advance came in The Minimalist Program, especially Chapter 4 of (1995). However, before this landmark there was significant development in his (1993: 172 ff.) in the form of Checking theory. As part of the drastic rethinking of linguistic theory at this time, the notion ‘Government’ was eliminated in favor of the independently motivated relations of X-bar theory (‘Spec-head’, ‘head–complement’, with ‘head–head (of complement)’ added. Case and agreement, now fall under Spec-head relations, with Case involving new Agr heads adapted from Pollock 1989. Verbal inflection is due to head-head adjunction for feature checking, where for example T may bear a V-feature and V bears an (uninterpretable) Infl-feature corresponding to each inflection that it bears. Under successful checking, the Infl feature is deleted; if it remains at LF the derivation crashes. ‘Raising’ vs ‘lowering’ is replaced by feature checking, either overt or covert which is constrained by having to fall under one of the X-bar relations. Choices as to overt or covert displacement are governed by the economy principle ‘Procrastinate’(1995: 314).

The next technical innovation was the development of ‘strong’ features (“one element of language variation” – 1995:232/5) and their use as triggers of overt phrasal and/or head-adjunction operations. The system of merge replaced X-bar structure with “Bare Phrase Structure”, under which a lexical item forms (part of) the label. That is, if α and β are merged the label of the resulting constituent must be either α or β (1995:242ff.). The notion ‘complement of a head’ is reduced to ‘first-merged phrase’; all other phrases merged (excluding adjuncts) are Specifiers of the head (1995:245). Under somewhat unnatural conditions, a head may permit one, two or more specifiers (1995:372ff, and 215). However, essential use of the term Specifier is still made (e.g. in defining the edge of a phase, in 2000: 108), so the notion Specifier has not yet properly been eliminated.6

In Chapter 4, there was a greater emphasis on formal features which contrast with semantic features such as [artifact]. Formal features are either interpretable at LF or not: [±interpretable]. Categorial features and some φ-features are taken to be interpretable, where uninterpretable features must be deleted (rendered invisible at the LF interface) for convergence (1995:277). Formal features of a lexical item were also either intrinsic or optional where an item selected from the lexicon is merged with its optional features specified— e.g. book, [accusative, plural]. If a head has optional features, these features must then be checked, by raising to some appropriate functional head or its Specifier under ‘Move F’ (1995:261ff.), if necessary along with further material required for convergence at PF (if nothing but a feature needs to be moved, we have ‘covert movement’). Only unchecked features move and the raising cannot skip intervening features of the right kind. That is, the unchecked features F move upwards, to seek

---

3 For reasons of space we presuppose considerable familiarity with the current scene.
4 We give page references from the version in Chapter 3 of 1995.
5 The first discussion of ‘weak’ vs. ‘strong’ elements is in Chapter 2:135ff.
6 See Cormack 1999 for arguments that it should be, with subjects and objects both merged before the verb.
some c-commanding target $K$ with which to match. $K$ is a legitimate target only if it too has an unchecked feature.

The most significant and far-reaching innovation for feature theory came with Phase theory (2000, 2001, 2007, 2008), with the claim that Move was a composite of (internal) Merge and Agree, the only processes of the grammar. All previous features and rules and other stipulations were subjected to scrutiny, and many eliminated (2000:132). Both Merge and Agree require a feature $F$ that drives the operation. For external Merge, this is a (semantic) selection feature, and the selector projects. For Agree, in 2000:127ff. the Probe bears an uninterpretable feature $F$ and searches downwards within its complement for the closest matching feature (an Active Goal). In 2001:5, uninterpretable features enter the derivation unvalued, so that Match is now defined by ‘non-distinctness’ as in 1965:84.7

The feature system is still somewhat informally indicated. In his 2001 Chomsky has applied it to some of the more recalcitrant problems in the literature, including Scandinavian transitive expletives and object shift and he has made further suggestions for parametrization without pursuing them in detail. A range of possible problems, especially regarding intervention effects, are also noted. Head-head adjunction is problematic for minimalist principles, leading Chomsky (2000 fn. 68) to suggest that it should be relegated to the Post-Spell-Out phonological component.

3 Features, Language Variation and Compositional Agree

After this outline of the feature theory implicit in Aspects and its development up to the mainstream Minimalist Program, we turn to a CCG take on Minimalism.

Suppose that the variation between I-languages, such as SVO vs. SOV, is determined lexically, under Merge, where lexical information includes features. Major variation between languages is probably due to the features of functional heads, and minor variation to features of classes of lexical heads, with micro-variation due to features of individual heads (counteracting the default of its lexical class).

The majority of linguists utilize features to explain structures or to formalize theories, but the properties and behavior of the features themselves is not always made explicit enough for the consequences to be clear. Further, many invoke additional principles or language-specific parameters. We want to show here how a fully merge-driven feature system of feature checking, together with a theory of possible feature structures, eliminates several stipulations and some problems remaining in the current Minimalist system.

We suggest a highly restricted compositional theory of features. In Cormack & Smith 2012, we proposed a ‘bottom up’ version of Agree, where the valuation of features on two items could take place only under Merge, hence, assuming binary merge, only under sisterhood of the items. This single system accounts for inflection and the displacement of ‘head movement’ (formalizing ‘Agree’), and extends with distinct features to phrasal displacement. It can also be used to account for selection restrictions that do not, or should not, fall under the categorial system.

Before elaborating further we need to introduce one or two presuppositions, in particular the relation of Natural Language (NL) and the Language of Thought (LoT), more precisely the

---

7 X is non-distinct from Y if the two do not disagree on any feature value. ‘Non-distinctness’ is the symmetric feature relation ultimately taken over from phonology.
relation of phrases in NL to phrases in LoT. Although specified in relation to CCG these presuppositions generalize, we believe, to all generative theories. We begin by discussing the syntactic background, and then the feature system, and then discuss three configurations for checking. For the first two of these, the only essential from the CCG background is that the LF is given, with a fixed order. Morphosyntactic features cannot change this; they can affect only the spell-out positions of certain morphosyntactic feature bundles (i.e. they can affect PF). For the final configuration of feature checking, it is necessary to exploit the flexibility given by the combinatorial system.

3.1 NL and LoT

Suppose that humans are equipped with a ‘language of thought’, whose syntax is based on arity, and type. Minimally, there are two types, <e> for entities and <t> for truth values. These allow the encoding of the arity of an LoT item. One-place and two-place predicates, for example, have types <e, t> and <e, <e, t>>, alternatively (e → t), (e → (e → t)), and so on. The rightmost type is the ‘goal’ type (the maximal mother type of a phrase headed by the item); the left hand types are those of items that may be merged, so that they serve one of the purposes of Chomsky’s (2008) Edge features. LoT lexical items of suitable types may be merged under function-argument application to give well-formed propositions (or smaller phrases) such that the meaning of the whole is a simple function of the meaning of the parts. This allows LoT to be used for inference.

The essential recursive step for a grammar of LoT that will serve for inference with IF, and simultaneously as the basis for NL, is that in (1), where P and Q are items of LoT, and α, β are types. P combined with Q in this order yields the result ‘P applied to Q’, with the types as shown. The arrow below indicates ‘…merge to form a category with the property …’

(1) item 1 item 2 mother
P Q → [P.Q]
<β, α> <β> <α>

If the selection type β for a lexical item P of LoT may be <t>, then recursion can occur in LoT, allowing for example thoughts about propositions, such as that corresponding to ‘The tiger does not know I am here’. Here, negation has type <t, t>, and the equivalent of know, type <t, <e, t>>. No further syntactic machinery is required, though the lexicon may be elaborated to include higher-type operators.

The first question then is why the syntax of NL is not as simple as this — why NLs are varied in syntax, not just in PF exponence of an LoT item. The second question is what minimal additional resources and appropriate constraints on them allow this — where the ‘minimal’

---

8 It is worth noting that we share with Chomsky the assumption that the emergence of a language of thought was evolutionarily prior to its externalization in the form of natural language (see Hauser et al, 2014:6), even though we are ignorant of “when such internal computations were externalized in spoken or signed language”.

9 The implicit priority this affords to LF is again consonant with Chomsky’s (2014) remark: “we should revise the Aristotelian picture of language as sound with meaning; rather, it should be regarded as meaning with sound, a very different conception.” In spirit the observation goes back to Aspects (p.16) where “for each sentence a deep structure … determines its semantic interpretation.” The prime example of such precedence appeared with the generative semanticists (e.g. Lakoff 1971).

10 Chomsky’s semantic features appear to have no classificatory syntax such as type; it is not clear whether they have a syntax, beyond (possibly) selection.
should account for why NLs are so similar over and above what is entailed by the common underlying LoT syntax. The Minimalist approach to NL, given LoT of this form, should be that LoT provides UG (the initial state) and processes permitting the acquisition of NL by an infant equipped with a suitable probabilistic inferential system.\footnote{See 2007:9 for skepticism, but accepting the LoT as an appropriate language for the CI interface. For the probabilistic inferential mechanism see e.g. the work of Xu and her associates.}

The answer to the first question should, as Chomsky argues (2000:120-121), be related to the demands of externalization. One facet of this relates to discourse considerations, such as the identification of topic, focus and so on by the displacement of phrases to designated positions. These processes aid communication by directing pragmatic processing even though they are putatively more costly than internal inferential processing. The morphosyntactic feature system is in part a simple solution to reducing the processing costs of externalization. A second facet of externalization is the requirement for rapid acquisition by the child. In this domain there is a plethora of evidence that by 18 months infants prefer grammatical sequences like \textit{is} ... \textit{walking} to ungrammatical sequences like \textit{can} ... \textit{walking} (Santelmann \& Jusczyk 1998), and that preverbal infants must have encoded bound functional morphemes by 11 months (Marquis \& Shi 2012). The morphosyntactic feature system provides a straightforward encapsulation of the relations that need to be acquired.

The classification involved is that of syntactic category. It seems likely that the categorial distinctions of NL aid both speaker and hearer in linking an LoT item to the output PF, as well as offering default inferential information (for example, that nouns relate to relatively permanent properties of objects, whereas adjectives tend to relate to relatively less permanent properties of those objects; or nouns and adjectives to states of objects, but verbs and adverbs to changes over time of those objects).

We take categorial features to include not only a goal category, such as N or V, but selection categories (corresponding to Chomsky’s subcategorisation features), such as D (the category for entities) or C (the category for declarative clauses).\footnote{For expository purposes we use a simplified system for both categories and types. Arguably, NL provides no instantiations for a D item, but for present purposes, a proper name may be taken to have that category.} We aim for a system where all items in the extended projection of V have goal category V, and so on for other lexical categories. The \textit{categorial feature system} includes well-formedness conditions: merge must respect these selection categories, as well as respecting semantic type selection. The natural assumption is that the two run in parallel, so that corresponding to a type $<e, <e, t>>$, there is a category $<D, <D, V>>$ for a transitive verb with goal category V and two selections for D (the category for entities). However, we use here the Combinatorial Categorial Grammar notation, so that $<D, <D, V>>$ is shown rather as (V/D)/D. The merge rule for categories runs in parallel to that for types, as in (2), so that no extra cognitive resources are required. Here, P and Q are items of NL, directly interpretable in LoT, and given external form.

\begin{align*}
\text{(2) item 1} & \quad \text{item 2} & \quad \text{mother} \\
\text{P} & \quad \text{Q} & \quad \rightarrow & \quad \text{[P.Q]} \\
\text{type:} & \quad \alpha/\beta & \quad \rightarrow & \quad \alpha \quad \text{(using the slash notation for types, here)} \\
\text{category} & \quad X/Y & \quad Y & \quad \rightarrow & \quad X
\end{align*}

Such a merge rule offers a natural linearization of items. Taking the left to right order on the page as corresponding to temporal order, the ordering can be described as ‘Functor First’
(‘Functor Last’ would also be possible). This linearization is foolproof: it gives a result for any input. It thus has two advantages over the proposal of Kayne 1994, and Chomsky’s 1995:335f. variant. First, it is indifferent as to whether the two items are drawn directly from the lexicon or not; one must always be the functor relative to the other, or they cannot merge. Second, adjuncts are unproblematic, since they obey the same rule as any other item (typically, they have category X/X for some host category X).

Combinatory Categorial Grammars extend expressive power by allowing more than one merge rule: one for each of the half-dozen Combinators permitted in the grammar. The required combinators include: I (identity, i.e. function application); B (function composition, allowing a ‘gap’ for the operand of some item, and hence for non-standard constituents); S (allowing a gap in each of the two operands of some item, e.g. in ‘across the board’ extraction), and arguably R (Cormack 2006) and Z (Jacobson 1999 and later), relating inter alia to control and binding respectively. We propose that the combinators are interpreted instead as lexical items, equivalent in essential respects to functional heads in Minimalism; then only one Merge rule is needed — function application, where this applies not only to standard lexical items, but to each of the permissible combinators when it is merged. As it is indeed only a subset of the possible combinators that is ever used in NL, placing these in the lexicon (as higher order items) is natural. This has the happy effect that the representation may be bracket-free — brackets are shown only for the reader’s convenience. A linear string represents unambiguously just the same information as a tree, which simplifies inference. We propose that neither the combinator C (which reverses the selection order of two operands of a head), nor any ‘backward’ combinators, occur in NL or LoT. This usefully restricts the choices to be made by the producer or the parser (and in particular, the child acquiring his native language).

The effect of function composition is produced by the combinator B, as in (3) (where the intermediate merge of items 1 and 2 is not shown):

\[
\begin{array}{c|c|c|c}
\text{item 1} & \text{item 2} & \text{item 3} & \text{mother} \\
B & P & Q & \\
\hline
\text{category:} & X/(Y/Z)/(X/Y) & X/Y & Y/Z & \rightarrow & X/Z \\
\text{type:} & (\alpha/\gamma)(w/\gamma)/(\alpha/w) & \alpha/\beta & \beta/\gamma & \rightarrow & \alpha/\gamma
\end{array}
\]

Because function composition equates \([\alpha [\beta \gamma]]\) it permits the non-standard constituents alluded to above, and exemplified in (4) and in (5):

\[
(4) \quad [B \text{ the first}] \text{ and } [B \text{ the third}] \text{ emperors were Greek}
\]

\[
(5) \quad \text{The third song, I thought that } [B \text{ no-one liked}]
\]

The combinators, including B, permit structures and meanings that would not be available if the only mode of combination were function-argument application.

We propose that any NL is built on such a basis, with the ordering Functor First at LF. Surface deviation from the CCG order is only superficial: some additional resource permits changes in order that affect PF, but not LF. The resource is the morphosyntactic feature system.

We will not pursue the combinatorial system for NL further here (for argument, examples, and extensions, see Cormack & Smith 2012, 2014, in prep.).

13 A further consequence is that the existence of strong islands is predicted — a matter we cannot pursue here.
3.2 Feature Checking Under Unification

In a feature system using unification, a feature is minimally some attribute together with its value, stated here as \(<\text{attribute}: \text{value}>, where attribute and value are also features. Attributes may be features like \text{NOUN}, \text{AUX}, or categories like D. Values may be the names of particular lexical items, like \text{DEVOUR}, or \text{PROG} (for the progressive), but may also be underdetermined (notation: \(u\)). Features may be isolated, but may also be interactive. What makes them interactive is that the value may be underdetermined (i.e. the feature is unvalued or partially valued). A feature which is valued may combine by unification with one that is undetermined, and thereby transmit information to the latter, giving it a determinate value. We assume that unification is given by the cognitive system, rather than UG, that is they are a function of ‘third factor’ considerations (2005). Features may also unify without full valuation (for example in some multiple agreement). But by their nature, such features on distinct items do not make any selections, so that unlike categories, they have no intrinsic structural way of combining. Instead, they may be taken to be parasitic on syntactic merge to license their interactions.\(^{14}\) The results of merge are given in (6). It is the third stipulation that allows unification and valuation of features of non-adjacent items, replacing Chomsky’s metaphor of the Probe ‘searching’ its c-command domain.

\(^{14}\) The standard claim is that there has to be a c-command relation with ‘search’.

\(^{15}\) Unification includes not only valuing the features under sisterhood, but transferring the values to the heads where the features originated. Thus the system includes the same search space for unification as Chomsky’s downward probing.

\[a\] If two items are syntactically merged as sisters, then feature-valuation of MSyn features on these items will occur by unification if possible.\(^{15}\)

\[b\] The result of the operation is recorded as a feature of the mother (as with selection features), and the unification values are passed down to the daughters via the intermediate nodes.

\[c\] An unvalued feature which cannot be valued under sisterhood percolates to the mother.

As well as these merge rules, we need conditions for failure:

\[7\] a Underdetermined features cause failure of the derivation

\[b\] If X and Y are merged, and X and Y have features with the same attributes, then the values of these features must unify at this merge — otherwise the derivation fails.

The condition in (7b), which we dub ‘Now or Never’, is the one responsible for ‘defective intervention’ effects (Chomsky 2000: 123 for features), to be illustrated below.

We will demonstrate that this system provides an adequate basis for a feature-based account of local non-categorial selection and verbal inflection. Elsewhere, we have shown how it can account for such phrasal displacement as does not affect LF (using category as the attribute, and PF as the value, for the relevant features).

We proposed that features came in two varieties:
The attributes here are word-classes. The values are the names of lexical items. As usual, a value may be undetermined, notated as ‘\(u\)’.

Under the merge rules given in (6) and (7), there are just three possible configurations for successful transfer of information from one head to another. These are ‘mutual checking’, ‘simple checking’, and ‘eavesdropping’. We illustrate these in turn. The first is essentially indifferent to configurational relations between the heads, but the other two impose tighter relations between one head and the other.

We first demonstrate with English inflection, arguing that the system eliminates any specification of upward vs. downward checking and search domains, captures relevant intervention effects without any Phase head being specified, and naturally accounts for head-movement.

### 3.3 Mutual Checking

For mutual checking, the two items bear diploid features such that each supplies a value which the other requires. In relation to the Probe-Goal checking of 2004 the first feature corresponds to the unvalued feature of the Probe, and the second to an unvalued feature of the Goal (required by the Activity condition). A nice example is given by the checking for inflections in the English verbal system, where examples like *He might have been being teased* are possible, but not all the heads involved are obligatorily present. Informally, certain heads require inflection; certain heads assign inflection. These have to be properly paired, across intervening items such as noun phrases and adjuncts. Diploid features make the required pairings, as we demonstrate in a moment. Given the predominantly right branching structure of English, it is clear that the goal needs to have an unvalued feature for percolation. But the probe too may be required to percolate if for example it is within an adjunction or coordination structure, as is *has* (twice) in (8), which does not c-command *eaten*.\(^{16}\)

(8) He either has or hasn’t eaten anything

The word-class of each inflecting item in the verbal projection (verbs, auxiliary, and modal items) unifies with V-EXT (“extended verbal projection”). The word-class of each item in the extended verbal projection which is capable of assigning inflection (modal, auxiliary, tense) unifies with AUX. Appropriate fully valued sets of features may be mapped to PF forms. In (9), the first feature in each diploid gives a verbal item requiring inflection, and the second, the source of the inflection.

(9) head 1. LF: GIVE, Morphosyntactic feature: \(<\text{VERB}: \text{GIVE}>, <\text{AUX}: u>>;\)

head 2. LF: PAST-SHIFT, Morphosyntactic features: \(<\text{V-EXT}: \text{u}>, <\text{AUX}: \text{PERFECT}}>\), \(<\text{AUX}: \text{PERFECT}>, <\text{AUX}: u>>\)

\(^{16}\) For an idea as to how both verbs can be checked, see Cormack and Smith 2005.
GIVE will be merged first. When the \textit{PERFECT} head is merged, the diploid from \textit{GIVE} will (after percolation if necessary) be sister to the features of \textit{PERFECT}. Unification can take place with the first of these features, since \texttt{\langle VERB: GIVE\rangle} unifies with \texttt{\langle V-EXT: u\rangle}, and \texttt{\langle AUX, u\rangle} unifies with \texttt{\langle AUX: PERFECT\rangle}. Then under unification, both heads and the mother will acquire the fully valued diploid feature \texttt{\langle\langle VERB: GIVE\rangle, \langle AUX: PERFECT\rangle\rangle}. We call this cooperative arrangement ‘Mutual Checking’. The underdetermined feature from \textit{PERFECT} will percolate until it finds itself in a sisterhood relation with the diploid feature from some other AUX head — say Tense.

Only the intervention of percolating features from another head with the same pair of attributes can forestall this unification (under ‘Now or Never’ in (7)). If the diploid from the verb finds itself in a sisterhood relation with an attribute-matched diploid from any other AUX item, then unification (or a crash) will take place then; conversely, if the underdetermined diploid from the AUX item finds itself sister to a feature-matched diploid from a V-EXT before that from \textit{GIVE}, there will be intervention. Intervention effects do not arise from any head such as an adjunct or a DP, because it will not bear a diploid with the same two attributes.

The diploid \texttt{\langle\langle VERB: GIVE\rangle, \langle AUX: PERFECT\rangle\rangle} may be spelled out as \textit{given}. As in Chomsky 2001:39, for a set of diploids that have unified as under (6b), only one will be spelled out at PF. Where possible, this is the first, presumably as a consequence of left to right Spell Out, combined with the effect of economizing on articulation. It is this spelling out of a diploid under a higher head that gives rise to what is known as ‘head movement’. Head movement then requires neither head-head adjunction, nor to be relegated to the PF component. No c-command relation needs to be stipulated (nor indeed should be), nor any target–head (1995) or probe–goal asymmetry (2000). A merge-based feature unification system necessarily produces a ‘Minimal Structural Distance’ effect in head movement (2013). The ‘activity condition’ of (2000) corresponds to the presence of an unvalued feature on the goal. Such a feature must be present if it is required to percolate, so no stipulation is required.

This sort of Agree is suited to inflection and phi-feature agreement between heads, and to other cases where if one head is present, the other must also appear. This however does not exhaust the syntactic uses of Agree.

### 3.4 Simple Checking
Suppose the relation between the heads is asymmetric, in that head 1 requires the presence of head 2, but not vice-versa. Mutual checking then is not an option. One such situation is the requirement that certain modals in English, such as \textit{SHOULD}, unlike say the modal NPI \textit{NEED}, cannot fall within the scope of Pol \textit{NEG}. It may be assumed that they always fall above Pol (the polarity head position, for positive polarity also). One way of ensuring this would be to give Pol a dedicated category, for which a high modal could select. However, this would incur costs. The polarity head would no longer have goal category V like other items in the extended projection of a verb. And an adverb like \textit{usually} may fall above or below Pol, so would need two categorizations (\textit{usually doesn’t; doesn’t usually}). Simple checking, slightly surprisingly, suffices to obtain the ordering. The idea is that a high modal launches a simple underdetermined feature \texttt{\langle POL: u\rangle}. If this unifies with its target — \texttt{\langle POL: NEG\rangle} or \texttt{\langle POL: POS\rangle} — then this gives the information that the polarity head was indeed present in the appropriate locality. If it fails to find any such target, the \texttt{\langle POL: u\rangle} will remain unvalued, and the derivation will crash. Thus such a feature may act as a filter, ruling out derivations where the target was not present within the area the percolating feature can check.
In the simplest possible instance of simple checking, the two heads concerned would be sisters, but this is not available with heads in the clausal spine when they are merged in a configuration like that in (10a). Sisterhood is ensured if the high modal is required (via a further feature) to merge using B, composing the modal and the negation as in (10b). One step of percolation of <POL: u> puts the checking feature into a sisterhood relation with its target (Bold print indicates features obtained under percolation or unification under sisterhood, omitting the effects of unification operating downwards). Note that the reverse order in (10c) will fail, so that the scope of the modal over the polarity head is ensured.

The minimal instantiation of (10b) is not the only possible configuration for success. The dotted line shows regions in which an adjunct could be merged without disrupting the checking. Because the merge must be with B, and the configuration in (10c) fails, it is not possible for the modal to incorrectly use the feature on a higher negation to value its <POL: u> feature. That is, the locality required is necessarily induced in Simple checking, and the proper LF scope is also automatically induced by the Simple feature.

The target item must c-command the checking head from the right, given Functor First. There is no ‘Activity condition’. In Cormack and Smith 2014, we exploit Simple checking to produce LF-preserving phrasal displacement. In a configuration like that in (10c), a combinator with a feature <X: u> for some category X may check an <X: α> feature at one of its operands. If α is the PF of the phrase of category X, this induces displacement of the PF content α to the combinator position. Typically, X will be the category of the operand, but it can sometimes be a
category more deeply embedded. A combinator with such a feature thus performs the function of an EPP feature on its first operand.

3.5 Checking by Eavesdropping

Finally, there is the Eavesdropping solution to obtaining information. Here, the information-seeking head launches a diploid, which intercepts and obtains information from the mutual checking diploids of some other pair of heads. For example, consider a situation where in a noun phrase, attributive adjectives show number agreement, where number comes from a separate head in the noun phrase. Then determiner and number can be in a mutual agreement relation, exploiting diploids of the form $<\text{DET: } \alpha, \text{NUM: } u>$, and $<\text{DET: } u, \text{NUM: } \beta>$, where $\alpha$ and $\beta$ are fixed by the heads in question, or similarly if the noun head and the number head are in a mutual agree relation. But no adjective need be present, so mutual checking for the adjective is not viable. Instead, the adjective can launch a feature $<\text{DET: } u, \text{NUM: } u>$ or $<\text{NOUN: } u, \text{NUM: } u>$ or perhaps $<\text{WORD: } u, \text{NUM: } u>$, and utilize the valued form of this to determine what number agreement it should bear at PF.

According to which choice is made, there will be different locality conditions for the possible merge positions of the adjective or adjective phrase. Suppose for instance the feature launched by the adjective is $<\text{NOUN: } u, \text{NUM: } u>$. Eavesdropping will be successful if the adjective has a common mother with either the NUM head, or (more probably) the NOUN head below the common mother of the NOUN and NUM heads. One successful configuration where NOUN and ADJ have a common mother is shown in (11). The features from noun and adjective are in a mutual checking relation (and a fortiori both active), with features percolating as shown by the bold lines. The adjective could be embedded in an AP, and the noun in an NP without disrupting the Agree (subject to there being no interveners). Sooner or later, the adjective’s checking feature will be in a sisterhood relation to the relevant feature percolating from the noun, where it can ‘eavesdrop’ to find the value of the NUM feature after the Num head has merged and its features unified with those on the noun. In (11), sisterhood is immediate, so that any number information from unification percolating down to the intermediate head from the mother will be transferred to the adjective as well as to the noun.

(11)
If the adjective is within an AP merged with the noun to the right, then checking would also be possible. This could occur under a covert conjunction interpretation of noun modification (Cormack and Breheny 1994, Cormack 1995). Without any stipulations, Eavesdropping allows multiple checking — of more than one adjunct to the noun, here. We see again that non-mutual feature checking, unlike mutual feature checking, can contribute to restricting merge positions, i.e. to locality conditions related to some head.

The configurations discussed in this and the previous section are the only ones in which non-mutual unvalued features can be valued. They may in an incorrect configuration cause the derivation to fail because they remain unvalued. Diploid features with matching attributes, by way of contrast, will always either succeed in mutual valuation, or fail under the ‘Now or Never’ specification.

In the system being sketched here, ‘Functional’ heads include Combinators. These serve to determine the semantics of merge (e.g. function application, function composition), giving also the category of the result. The same feature system that provides for Agree can induce ‘displacement’ at PF to a combinator position, by exploiting category and PF as attributes in a diploid feature. Major word order variation (not affecting LF) can be encoded by setting such checking features on combinators that make particular categorial selections (Cormack and Smith 2014).

4 Conclusions

To conclude, we list the major similarities and differences between our proposals and recent work by Chomsky:

Like Chomsky we accord synchronic priority to LF, and evolutionary priority to LoT. Like him we use underdetermined features (with unification, implicit in Chomsky’s papers), and encode parameters as features on functional heads.

Unlike Chomsky we appeal to Categories (with categorial selection), our version of Merge operates ‘Functor First’ and, crucially, Combinators (as functional heads) constitute lexical items which play a role in the syntax. Our features can range over category as well as items from the lexicon, and we make extensive use of Diploid features. Agree is regulated only by Merge rules, and Agree + Spell accounts for LF-preserving head movement as well as for LF-preserving phrasal displacement. Finally, we claim that microparameters can be encoded on lexical heads.

We hope to have shown that a Feature System governed just by the five rules stated informally in (6) and (7) is worth exploring. It subsumes much of Chomsky’s 2004 Probe–Goal system, but also makes useful but tightly constrained predictions outside this. The radical difference is that with CCG combinators in the grammar, LF can be produced by categorial Merge, and feature-based displacement can be confined to PF effects — inter alia rendering the task of both acquisition and linguistic analysis simpler.

Aspects continues to inspire.

References


Cormack, Annabel and Neil Smith. 2014. Phrasal Displacement Within the Clause. lingbuzz/001992


MENTALISM AND UNIVERSALS IN ASPECTS

TIM STOWELL
UCLA

The field of generative-transformational grammar emerged in the late 1950s as the ascendant research paradigm of modern linguistics, created and led by Noam Chomsky. When Aspects was published in August 1965, the field was already ten years old; Chomsky had already published two previous books on linguistic theory, and the first generation of his graduate Ph.D. students had completed their dissertations at MIT.

Nevertheless Aspects is effectively the most important foundational document of the field. First, Aspects provided the definitive exposition of the classical theory of transformational-generative grammar—the so-called Standard Theory. It introduced influential constructs, such as the level of Deep Structure, and provided an outline of the theory of the Lexicon that has been quite enduring (though various theoretical frameworks have added to it in various ways). Nevertheless, the importance of Aspects rests less on its exposition of the technical details of the Standard Theory than on its presentation of Chomsky’s vision of linguistic theory as a mentalistic theory, a component part of a broader theory of the human mind. This position is set forth explicitly in the first section of the first chapter (p. 4):

The problem for the linguist, as well for the child learning the language, is to determine from the data of performance the underlying system of rules that has been mastered by the speaker-hearer and that he puts to use in actual performance. Hence, in the technical sense, linguistic theory is mentalistic, since it is concerned with discovering a mental reality underlying actual behavior.

Theories of grammar, by definition, are “psychologically real” insofar as they seek to characterize properties of the system of unconscious knowledge that underlies human linguistic ability. This mentalistic interpretation of linguistic theory has been a central guiding assumption of mainstream syntactic theory ever since, even though it has never been universally accepted by linguists of all persuasions. It alone has been responsible for shifting the place of the field within the intellectual landscape, away from being an adjunct to Anthropology and into a central role within the Cognitive Sciences.
By viewing grammar as an attribute of the human mind, Chomsky provided the conceptual foundation for viewing linguistics as a natural science: in this conception, theories of grammar are no more and no less than theories about the nature of something that exists in the physical world—namely, the human mind/brain. Viewing grammar as a mental object also explains why native speakers’ intuitions about grammaticality and meaning are of central importance. This clarity of vision swept away numerous conceptual and metaphysical problems associated with non-mentalistic theories of grammar, such as the notion of a grammar as a “social construct.”

This mentalistic interpretation of linguistic theory gave new importance to the study of language acquisition by children. If the system of linguistic knowledge in the mind of a mature speaker-hearer is the field’s central object of study, then the way in which this system arises is of central concern too. Knowledge of grammar must be shaped during childhood, as a result of the child’s innate cognitive capacities responding to language data that it is exposed to in observing and participating in actual language use. Therefore the grammar of a particular language is the product of both the innate predispositions of the mind of the language learner and the sample of utterances in the particular language that the child is exposed to. The study of various intermediate stages of language acquisition may therefore inform, and be informed by, the general theory of grammar.

Neither The Logical Structure of Linguistic Theory (LSLT), completed in 1955, nor Syntactic Structures (1957) explicitly endorsed mentalism *per se*. Parts of LSLT actually seem to dismiss the usefulness of mentalistic constructs in grammatical theory construction, while Syntactic Structures takes a quasi-mathematical set-theoretic approach to language (p. 13):

I will consider a language to be a set [...] of sentences [...] The grammar of [a language] L will thus be a device that generates all of the grammatical sequences of L and none of the ungrammatical ones.

Both studies also seem relatively unconcerned with accounting for language acquisition by children. A more central concern is the question of whether, and in what way, general linguistic theory should guide a linguist who is seeking to discover the correct grammar of a language. In Syntactic Structures, Chomsky argued against a tenet of the Structuralist tradition, which held that linguistic theory should provide a set of discovery procedures by which a linguist can programmatically infer the correct grammar on the basis of exposure to a representative sample of utterances. He argued instead (p. 56) that the theory should simply provide a means to evaluate hypothesized grammars relative to each other:

we shall never consider the question of how one might have arrived at the grammar whose simplicity is being determined; e.g. how one might have discovered the analysis of the verb phrase presented in § 5.3. Questions of this sort are not relevant to the program of research that we have outlined above [...] Our ultimate aim is to provide an objective, non-intuitive way to evaluate a grammar once presented [...] We are thus interested in describing the form of grammars (equivalently, the nature of linguistic structure) [...] rather than in showing how, in principle, one might have arrived at the grammar of a language.

Tellingly, the concern here is with the linguist, rather than with the child. Indeed, children (unlike linguists) presumably *are* endowed with some kind of discovery procedure which guides them to a particular (and correct) grammar.
In his highly influential review of B.F. Skinner’s book *Verbal Behavior*, Chomsky (1959) endorsed an explicitly mentalistic interpretation of grammatical theory, positing it as the basis of an alternative to Skinner’s behavioristic account of language. Here the impressive feat of language acquisition by children is a matter of central concern for the first time:

The construction of a grammar [...] characterizes abstractly the ability of one who has mastered the language to distinguish sentences from nonsentences, to understand new sentences (in part), to note certain ambiguities, etc. [...] The speaker and the listener, we must assume, have already acquired the capacities characterized abstractly by the grammar [...] The child who learns a language has in some sense constructed the grammar for himself; [...] this grammar is of an extremely complex and abstract character [...] the young child has succeeded in carrying out what from the formal point of view, at least, seems to be a remarkable type of theory construction [...] The fact that all normal children acquire essentially comparable grammars of great complexity with remarkable rapidity suggests that human beings are somehow specially designed to do this, with data-handling or ‘hypothesis-formulating’ ability of unknown character and complexity.

The shift to mentalism is complete in *Aspects*; this new vision of the field, hinted at in the Skinner review, is presented in considerable detail in the first chapter. Many of the central issues associated with a mentalistic linguistic theory are identified and clarified here, including the crucial distinction between competence and performance. The problem of explaining language acquisition by children is discussed in depth, and the notion of an innate “language acquisition device” is introduced. The evaluation procedure that chooses between grammars is no longer just a theoretical aid to a linguist, but rather a facet of the innate acquisition device:

The device would [...] select one of these potential grammars by the evaluation measure [...] the device has now constructed a theory of the language of which the primary linguistic data are a sample. The theory that the device has now selected and internally represented specifies its tacit competence, its knowledge of the language. The child who acquires a language in this way of course knows a great deal more than he has “learned.” His knowledge of the language, as this is determined by his internalized grammar, goes far beyond the presented primary linguistic data and is in no sense an “inductive generalization” from these data. (pp. 32-33)

Moreover, theories of grammar are now judged in terms of their ability to account for language acquisition (p.27):

The problem of internal justification—of explanatory adequacy—is essentially the problem of constructing a theory of language acquisition, an account of the specific innate abilities that make this achievement possible.

In this context, Chomsky discusses the notion of linguistic “universals” and gives it a distinctly mentalistic and theoretical spin. He argues (pp. 27-8) that:

A theory of linguistic structure that aims for explanatory adequacy incorporates an account of linguistic universals, and it attributes tacit knowledge of these universals to the child. It proposes, then, that the child approaches the data with the presumption that they are drawn from a language of a certain antecedently well-defined type, his problem being to determine which of the (humanly) possible languages is that of the community in which he is placed.
Consequently, the main task of linguistic theory must be to develop an account of linguistic universals that, on the one hand, will not be falsified by the actual diversity of languages and, on the other, will be sufficiently rich and explicit to account for the rapidity and uniformity of language learning, and the remarkable complexity and range of the generative grammars that are the product of language learning.

Later on, he dismisses almost in passing Joseph Greenberg’s proposals regarding implicational universals of surface word order patterns (p. 136):

Insofar as attention is restricted to surface structures, the most that can be expected is the discovery of statistical tendencies, such as those presented by Greenberg (1963).

Chomsky posits two types of linguistic universals in *Aspects*: substantive and formal. Substantive universals claim that “items of a particular kind in any language must be drawn from a fixed class of items;” he cites Jakobson’s inventory of distinctive features and traditional universal grammar's inventory of the syntactic categories *noun* and *verb* as examples. On the other hand, formal universals “involve rather the character of the rules that appear in grammars and the ways in which they can be interconnected.” (p. 29) These are "universal properties of a more abstract sort:” (pp. 29-30)

Consider a claim that the grammar of every language meets certain specified formal conditions. The truth of this hypothesis would not in itself imply that any particular rule must appear in all or even in any two grammars. The property of having a grammar meeting a certain abstract condition might be called a formal linguistic universal, if shown to be a general property of natural languages. Recent attempts to specify the abstract conditions that a generative grammar must meet have produced a variety of proposals concerning formal universals, in this sense. For example, consider the proposal that the syntactic component of a grammar must contain transformational rules (these being operations of a highly special kind) mapping semantically interpreted deep structures into phonetically interpreted surface structures, or the proposal that the phonological component of a grammar consists of a sequence of rules, a subset of which may apply cyclically to successively more dominant constituents of the surface structure (a transformational cycle, in the sense of much recent work on phonology). [...] The existence of deep-seated formal universals, in the sense suggested by such examples as these, implies that all languages are cut to the same pattern, but does not imply that there is any point by point correspondence between particular languages... whereas substantive universals have been the traditional concern of general linguistic theory, investigations of the abstract conditions that must be satisfied by any generative grammar have been undertaken only quite recently. They seem to offer extremely rich and varied possibilities for study in all aspects of grammar.

On this view, linguistic universals are, fundamentally, innate properties of the mind; more specifically, they are properties of the mind's language acquisition device and also of the mature grammar of a speaker/hearer. They include constraints on the form and application of grammatical rules and processes.

Like almost every other proposal in the first chapter of *Aspects*, the conception of formal linguistic universals laid out there is still completely applicable today. Indeed, formal universals have largely eclipsed substantive universals as the standard examples of linguistic universals that generative syntacticians are inclined to cite today. On the other hand, the actual examples of
formal universals that linguists today are most likely to cite are far more intricate than the examples that Chomsky cited in *Aspects*. The formal universals that contemporary syntacticians are likely to cite include islands and locality constraints on movement, binding conditions on anaphora, licit and illicit patterns of agreement, principles governing the interface between syntax and logical form, and highly articulated principles of phrase structure of the sort proposed within so-called cartographic theories of functional categories. These changes reflect the natural development and further articulation of syntactic theory over the past half century.

It is striking that Chomsky was able to foresee the general character of these subsequent developments in his discussion of the issue in 1965, given that the available evidence for formal universals was comparatively scanty at that time. Very few languages other than English had so far been analyzed in depth within the framework of generative grammar; thus, the conjecture that the grammars of all languages contain fundamentally the same kinds of rules and constraints as are found in the grammar of English was largely an article of faith.

The two central rule systems of the *Aspects* framework were the set of context-free phrase structure rules of the base (discussed in Chapter 2) and the set of transformational rules mapping deep structures to surface structures (discussed in Chapter 3). Both rule types were largely unconstrained by current standards, thus leading to considerable uncertainty about the nature of actual grammars—the theory allowed for a lot of competing analyses of individual constructions. Moreover, it wasn’t even clear that all languages had rules of the type posited for English. Traditional analyses of syntactic cross-linguistic variation had bequeathed to the field a widespread assumption that many languages have “free word order,” leading many linguists to suspect that phrase structure rules might be unmotivated and unnecessary for such languages; a common view was that the grammatical functions served by syntactic phrase structure in languages with fixed word order like English were handled by rules of morphology and agreement in certain other languages. Even if the grammars of all languages that had been analyzed at that point were assumed to have transformational rules, there was still considerable variation in the actual rules involved, many of which were construction-specific. Constraints on rules were generally stated on a rule-by-rule basis. Thus, the prospects of finding much in the way of formal universals in these rule systems must have struck many people as being rather unlikely.

Thus, in proposing the theory of formal universals in *Aspects*, Chomsky relied more on the force of the logic of his account of grammar as a mental construct and his proposed solution to the associated problem of accounting for language acquisition. Not for the first or last time in his career, Chomsky had identified a problem that few other linguists acknowledged as a problem, and proposed a completely original solution to it, bringing with it numerous implications inside and outside of the field of syntax. There wasn’t a lot of substantive evidence supporting this view, but he had faith that such evidence would some day be found, and he was proved right.

References

ON BY-PASSIVE AND WHY UNACCUSATIVES DON’T PASSIVIZE

TARALD TARALDSEN
University of Tromsø

1 The Passive Transformation in Aspects

In Aspects (p. 104), Chomsky proposes a passive transformation with the structural description in (1):

(1) NP - Aux - V - … - NP - … - by passive - …

Chomsky takes by passive to be a manner adverb, and its presence in (1) is intended to prevent the passive transformation from applying to verbs that do not take manner adverbs, like weigh in The pumpkin weighed 100 kilos or resemble in He resembles his mother. Since weigh and resemble have non-agentive subjects in such sentences, it is natural to assume that by passive is intended to be an agent-oriented adverb in (1). Since unaccusative verbs like fall etc. do not have agentive subjects either, they too should then not be able to undergo a transformation with the structural description in (1). If the appearance of a dedicated passive auxiliary is also due to the passive transformation, this might account for the ungrammaticality of (2)b in Norwegian:

(2)a Huset ble bygget i 1645 (av en bryggerieier)
   the house became built in 1645 (by a brewery owner)
   “The house was built in 1645 (by a brewery owner).”

b*Katten ble forsvunnet i forrige uke
   the cat became disappeared last week

Of course, one might say that the structure underlying (2)b would fail to meet the structural description in (1) even without by passive, since an unaccusative verb only takes a single NP argument. But if the Aspects account of passives is imported into a framework where the second
NP in (1) must move into an empty subject position, as in Emonds (1976), and the sole argument of an unaccusative verb originates in the VP\textsuperscript{1}, the passive transformation would have a structural description matching the structure underlying *Katten forsvant* “The cat disappeared” but for the presence of *by passive.*

Equivalently, one might set up things so that passives can only be formed from verbs associated with “little v”, taking this to be the head that introduces agentive external arguments.

In this squib, I will show that that none of this will be quite sufficient to account for the contrast between transitive and unaccusative verbs exemplified in (2), and I will suggest that (2)b can only be excluded if one makes specific assumption about the way lexical roots are associated with syntactic heads.

## 2 Lundquist’s Observation

The crucial observation is due to Björn Lundquist (2008), who first points out that *bli* “become”, which is used as the passive auxiliary in Scandinavian (cf. (3)), also combines with adjectives, as in (3)b:

\begin{enumerate}
\item[(2)a] Huset \textbf{ble} bygget i 1645 (av en bryggerieier)
\item[(3)a] Katten \textbf{er} (fortsatt) syk
\item[(2)b] Katten \textbf{ble} syk i forrige uke
\end{enumerate}

“the house became built in 1645 (by a brewery owner)"

“the cat became sick last week”

Then, he notes that although adjectival participles (with the properties of Kratzer’s (2000) “target state participles”) can be formed from unaccusative roots, e.g. *forsvunnet* “disappeared” from *forsvinne* “disappear”, such adjectival participles cannot combine with *bli* “become”:

\begin{enumerate}
\item[(4)a] Katten \textbf{er} (fortsatt) forsvunnet
\item[(2)b] Katten \textbf{ble} forsvunnet i forrige uke
\end{enumerate}

\begin{enumerate}
\item[(3)b] Katten \textbf{ble} syk i forrige uke
\end{enumerate}

\begin{enumerate}
\item[(4)b] Katten \textbf{ble} forsvunnet i forrige uke
\end{enumerate}

\begin{enumerate}
\item[(3)b] Katten \textbf{ble} forsvunnet i forrige uke
\end{enumerate}

\begin{enumerate}
\item[(4)b] Katten \textbf{ble} forsvunnet i forrige uke
\end{enumerate}

\begin{enumerate}
\item[(3)b] Katten \textbf{ble} forsvunnet i forrige uke
\end{enumerate}

This clearly means that no formulation of the passive transformation would suffice to rule out (2)b, since there is an alternative derivational path to (2)b which must also be closed.

Equally clearly, we cannot prevent *bli* from embedding adjectival passives from unaccusative verbs by making it select for vP, since the adjective in (3)b is certainly not a vP.

\textsuperscript{1} Actually, pseudo-passives like (i) suggest that the argument one generally takes to originate as an object may be replaced by an agentive DP under certain circumstances, especially on Collins’s (2005) analysis of *by*-phrases:

(i) \hspace{1cm} Penn Station is arrived at by 2000 commuters every morning
Lundquist’s own conjecture is that (4)b somehow is blocked by the existence of (5) with the same meaning as (4)b would have, i.e. transition into a state:

(5) Katten forsvant i forrige uke
the cat disappeared last week

In the following sections, I’ll sketch a way of implementing this idea.

3 Decomposing the Verb

Ramchand (2008) argues for the following decomposition of verbs:

(6) \[[Init DP_1 Init [Proc DP_2 Proc [Res DP_3 Res]]]]

In (6), the ResP denotes a state holding of the individual(s) denoted by DP_3, whereas ProcP denotes a process in which the individual(s) denoted by DP_2 (“the undergoer”) participates. DP_2 and DP_3 may be the same DP, i.e. Spec-ProcP may be created by Internal Merge. Init introduces an “initiator” (DP_1), i.e. an individual in a state that triggers the process denoted by ProcP, and we shall take it that DP_1 may be the same as DP_2. (This will become crucial in section 6.) Thus, Init is a near-equivalent of the head called “little v” (or, perhaps, v*) in other accounts.

The subeventualities denoted by the three different heads are tied together by the “leads-to relation”. For example, the process denoted by Proc leads to the state described by the ResP. In the case where DP_2 = DP_3, we may say that the process leads the undergoer into a certain state.

From this perspective, we may fit (3)b into (6) by taken bli “become” to lexicalize Proc, since (3)b describes the cat’s transition into a state of illness:

(3)b Katten bli syk i forrige uke
the cat became sick last week
“The cat got sick last week.”


This resonates with Baker’s (2005) idea that a verb may contain an adjective at the bottom of its internal structure.

Ramchand also argues that a single verb may lexicalize more than one single head in (6), depending on its lexical properties. So if (the root of) forsvinne “disappear” is specified in the lexicon as able to lexicalize both Proc and Res, we get the picture in (8):

---

2 According to Ramchand (2008), Init and Res are optional components of the template. That is, a verb may only have Proc in it.

3 For present purposes, we may think of this as an instance of “spanning”: If the lexical entry associates a morpheme (lexical or functional) with sequence of heads S, e.g. <Init, Proc, Res>, that morpheme can lexicalize any sequence S’ of hierarchically adjacent heads (adjacent modulo Specifiers) if and only if S’ matches a subsequence of S. (This is an adaptation of Starke’s (2009) Superset Principle.)
If the lexicalization algorithm prefers to use the root lexicalizing Res to lexicalize higher heads as well, whenever the lexical specification of the root allows this, we can capture the blocking effect assumed in section 2, i.e. (9) is excluded:

(9)*[[Proc the cat Proc [Res <the cat> Res ]]]

| bli          | forsvinne |

(3)b is fine, because syk and other adjectival roots cannot lexicalize heads above Res, but (4)b is blocked by (5), since forsvinne can lexicalize heads above Res.

We still allow for (4)a by assuming that the copula være “be” is not the spell-out of any part of (6), unlike bli “become”.

In the next section, I’ll show how this can be made consistent with the appearance of bli with participles formed from (agentive) transitive verbs.

### 4 Lexicalizing the Decomposed Verb

The problem we need to solve is this: Given that (10) shows that the root of bygge “build” can lexicalize all three heads of (6), as in (11), we might expect (2)a to be blocked for the same reason as (2)b/(4)b:

(10) Bryggerieieren bygget huset i 1645.

the brewery owner built house-the in 1645

(11) [[InitP the brewery owner Init [Proc the house Proc [Res <the house> Res ]]]

|__________________________|

bygge

(2)a Huset ble bygget i 1645 (av en bryggerieier)
the house became built in 1645 (by a brewery owner)
“The house was built in 1645 (by a brewery owner).”

(12)a [[Proc the house Proc [Res <the house> Res ]]]

|__________________________|

bygge

---

4 This should be seen as a consequence of (an adaptation of) a principle that nanosyntacticians variously refer to as “Biggest wins” or “Minimize junk”.

b*[Proc the house Proc [Res <the house> Res ]]
  bli bygge

Having *bygge* block *bli* as in (12) in fact gives the right result for adjectival participles of the type called “resultant state participles” by Kratzer (2000), i.e. the adjectival participles involving transition into a state:

\[
\text{(13) a) Huset \ antas \ bygget \ i \ 1645} \\
\text{the house \ is.assumed \ built \ in \ 1645} \\
\text{“The house is assumed to have been built in 1645.”} \\
\text{b*Huset \ antas \ bli \ bygget \ i \ 1645} \\
\text{the house \ is.assumed \ become \ built \ in \ 1645}
\]

But (2)a must obviously not be blocked. What I propose, is that verbal passive participles also include Init, and that *bli* “become” lexicalizes Init in (2)b:

\[
\text{(14) [Init Init [Proc the house Proc [Res <the house> Res ]]]} \\
  bli bygge
\]

(The assumption that verbal passive participles contain Init, is based on taking *by*-phrases to link a DP to Init.)

This, of course, also contradicts earlier assumptions, since we have seen that the root of *bygge* must be able to lexicalize Init in addition to Proc and Res in active sentences like (10). To fix this, we’ll have to distinguish between two “flavors” of Init. One, notated as \(\text{Init}_A\), will only appear in active sentences, while the other, \(\text{Init}_P\), only appears in passives.\(^5\) We can then say that the sequence of heads that can be lexicalized by the root of *bygge* and other transitive verbs includes \(\text{Init}_A\), but not \(\text{Init}_P\), e.g. the lexicon associates the root of *bygge* with the sequence \(<\text{Init}_A, \text{Proc}, \text{Res}>\), and the general lexicalization algorithm will allow it to lexicalize any subsequence of that sequence, but nothing else. This will prevent the root of *bygge* from blocking *bli* in (15):

\[
\text{(15) [Init Init [Proc the house Proc [Res <the house> Res ]]]} \\
  bli bygge
\]

This presupposes that *bli* is associated with the sequence \(<\text{Init}_P, \text{Proc}>\). Only in this sense is *bli* a passive auxiliary. Since *bli* can lexicalize any subsequence of \(<\text{Init}_P, \text{Proc}>\), it will not only

\(^5\) This begs the question how exactly the difference between \(\text{Init}_A\) and \(\text{Init}_P\) is to be characterized. This seems equivalent to characterizing the difference between different “flavors” of little \(v\) or of Voice. I leave the question open here.
appear in passives, but also in sentences like (3)b, which most likely don’t contain any Init in their structure:

(3)b  Katten ble  syk i forrige uke
    cat-the became sick last week
    “The cat got sick last week.”

As for the structures in which an unaccusative verb like forsvinne lexicalizes Proc and Res, either they cannot also contain Init or they can only contain Initp in which case these unaccusative verbs will be treated as a sort of lexicalized passives. Either way, there will not be room for bli. But to the extent that unaccusative verbs don’t come with agentive by-phrases, the first of the two options may be preferable.

5 Participles vs. Non-Participial Forms

In (5), where the verb root lexicalizes Proc and Res, the verb ends up in a non-participial form:

(5)  Katten forsvant i forrige uke
    the cat disappeared last week
    “The cat disappeared last week.”

But in (2)a and (13)a, where the verb root also lexicalizes Proc and Res, the verb must come out as a participle:

(2)a  Huset ble bygget i 1645 (av en bryggerieier)
    the house became built in 1645 (by a brewery owner)
    “The house was built in 1645 (by a brewery owner).”

(13)a  Huset antas bygget i 1645
    the house is assumed built in 1645
    “The house is assumed to have been built in 1645.”

Even forsvinne must surface as a participle when it only lexicalizes Res:

(4)a  Katten er (fortsatt) forsunnet
    cat-the is (still) disappeared

---

6 This is not entirely obvious, however. To the extent that (3)b may be taken to describe an eventuality in which the cat found itself in a state that led to a process of its getting sick, (3)b might well contain Init. But then bli or the adjectival root lexicalizing Res must somehow constrain Init to fill its Spec by movement of the undergoer DP in Spec-ProcP (see the discussion of smelte “melt” in section 6) so that (i) is excluded:

(i)  *Maten ble katten syk
    the food became the cat sick

If (3)b contains InitA, the sequence of heads associated with bli must contain an Init which is unspecified for the properties distinguishing InitA from Initp. This would not allow bli to occur in active sentences with a verbal root lexicalizing Proc and Res, since the root would block bli at InitA anyway.
In terms of the assumptions I have made along the way, there is a descriptive generalization that can be formulated as in (16):

(16) A verb must be(come) a participle when it fails to lexicalize all the heads in the sequence the lexicon associates its root with.

If the root of *forsvinne* is associated with < Proc, Res>, it will be packed into a participle when it only lexicalizes Res, as in (4)a, but not when it lexicalizes both Proc and Res, as in (5). If transitive roots like the root of *bygge* are associated with < Init<sub>A</sub>, Proc, Res>, they will surface as participles in passives, where they don’t lexicalize Init<sub>A</sub>.

However, a formal implementation of (16) would involve taking a stand on a number of issues that I’m not in a position to address here.

## 6 Anticausatives

The account sketched here has consequences for the analysis of verbs that appear both as unaccusative verbs and as agentive transitive verbs, like *smelte* “melt” in (17) which exemplifies a “causative/anticausative alternation”:

(17)a  Vi smelter isen
       we melt    the ice
b  Isen smelter
    the ice melts
c  Isen ble smeltet
    the ice became melted
    “The ice was melted.”

As (17)b shows, such verbs occur as participles combining with *bli*. If (16) is a valid generalization, this is inconsistent with a particular way of relating (17)a and (17)b. Saying that a causative/anticausative pair corresponds to a single root which can lexicalize all three heads in (18)a as well as only the two heads in (18)b, is consistent with associating the root of *smelte* with the sequence < Init<sub>A</sub>, Proc, Res >, since a root is by assumption capable of lexicalizing any subsequence of the sequence it is associated with:

(18)a  [InitP vi Init [ProcP isen Proc [ResP <isen> Res ]]]  ( = (18)a)
        |___________|______________|
        |___________|
        smelte

(18)b  [ProcP isen Proc [ResP <isen> Res ]]  ( = (18)b)
        |___________|
        |___________|
        smelte
But if (16) is valid, the verb should then be a participle in (17)b where one of the heads in <Init,
Proc, Res> is not lexicalized by the root of smelte.

We might try to get around this by marking Init as optional in the sequence of heads associated with smelte, and by modifying (16) so that it applies only when a root fails to lexicalize all non-optional heads in the sequence it is associated with. But then the root should not surface as a participle in (17)c, where the two non-optional heads in <(Init, ) Proc, Res> are both lexicalized by smelte:

(19) \[InitP \text{Init} \text{[ProcP isen Proc [ResP <isen> Res ]]}\]

\[\text{bli} \quad \text{smelte}\]

Therefore, if (16) is valid, we must adopt a view of anticausatives similar to the one advocated by Chierchia (2005) and analyze (17)b as in (20), where the event described by (17)b is represented as self-initiated, i.e. the ice is initially in a state (e.g. of having reached a certain temperature) which sets off a process of melting that terminates with the ice being melted:

(20) \[\text{Init isen Init}_A \text{[ProcP <isen> Proc [ResP <isen> Res ]]}\]

\[\text{smelte}\]

In (20), all the heads in <Init, Proc, Res> are lexicalized by the root of smelte, and (16) does not predict that smelte should appear in a participial form.

But in (19), the root doesn’t lexicalize Init, and must therefore show up as a participle.

7 Conclusion

The discussion extending over the preceding sections started out from two observations. The first was that the inclusion of by passive in the structural description of Chomsky’s passive transformation in Aspects could be seen as a way of limiting the passive transformation to agentive verbs, thus excluding unaccusative verbs as well as non-agentive weigh and resemble.

The second observation was the observation that the “passive auxiliary” bli “become” also can embed adjectives, as in (3)b:

(3)a Katten er (fortsatt) syk
the cat is (still) sick

b Katten ble syk i forrige uke
the cat became sick last week
“The cat got sick last week.”

Given the fact that some unaccusative verbs can form adjectival participles, as in (4)a, which seems parallel to (3)a, this leads to the expectation that adjectival participles can also be
embedded under bli giving rise to sentences like (4)b/(2)b, which would then be entirely parallel to (3)b:

(4)a  Katten er (fortsatt) forsvunnet
       the cat is (still) disappeared
b*Katten ble forsvunnet i forrige uke
       the cat became disappeared (last week)

Since (3)b clearly isn’t created by anybody’s passive transformation, this means that what we would expect, is that there should be an alternative derivational path leading to exactly the same set of surface strings that would be created, if the passive transformation were allowed to apply to unaccusatives. But as Lundquist has pointed out this path must be closed, since (4)b is in fact ungrammatical. The rest of this squib has been devoted to discussing a particular way of achieving this.

In conclusion, it should be pointed out that the account I have sketched, still retains a residue of by passive insofar as the presence of an agent-oriented manner adverb by passive in the passive transformation proposed in Aspects should be seen as a way of enforcing the presence of a little v introducing agentive external arguments and little v would be equivalent to the head Init appearing in (6). Crucially, the account of (4)b had to be set up so that bli can combine with a participle only in structures that contain Init (in its passive guise).

References

1 Introduction

The quote above reveals that Chomsky had minimalist thoughts already in 1964. In (2005) he attributed the so-called Strong Minimalist Thesis to “principles of structural architecture and developmental constraints that enter into canalization, organic form, and action”. It is useful to reflect on what it would take to “ground in physical law” some of the properties of linguistic systems that go back in their initial formulation to *Aspects*.

2 Quantum Cognition

If contemporary physics is to help understand a linguistic computation, the apparatus of quantum theory, dealing with distributed systems, may well be a starting point. Generative linguists would not be the first to initiate such efforts: the field of quantum cognition, for staters, already exists, with reasonable questions and challenges. Their Wikipedia entry starts as follows:

Quantum cognition [. . .] applies the mathematical formalism of quantum theory to model cognitive phenomena such as information processing by the human brain, decision making, human memory, concepts and conceptual reasoning, human judgment, and perception.

The field is not to be confused with any hypothesis that the brain itself involves quantum processes. While quantum cognition is compatible with such a possibility, the theory is more
abstract. It is based on the idea, expressed in Aerts and Aerts 1994, Atmanspacher et al. 2002, Khrennikov 2006, that quantum probability theory can help understand information processing in minds, by factoring in probabilistic notions and contextual dependencies. More specifically, quantum cognition:

\[ \text{... uses the mathematical formalism of quantum theory to inspire and formalize models of cognition [focusing] on modeling phenomena in cognitive science that have resisted traditional techniques or where traditional models seem to have reached a barrier (e.g., human memory), and modeling preferences in decision theory that seem paradoxical from a traditional rational point of view (e.g., preference reversals). [Wikipedia]} \]

The implication is that neural networks produce effects that can be modeled by the interference of probability amplitudes (a wave which, upon measurement, gives rise to a probability).

To get a sense of that mouthful non-technically, consider tuning a guitar. This involves settling on the high E (string #1), then proceeding to equalize the remaining strings in pitch at a higher fret. At stake is the frequency at which a string vibrates when plucked, as altered by tensing the tuning keys. One knows two strings have been equalized by successively listening to the notes each plays: the shorter one listens, the less clear the analysis. The question is why a very short (i.e. very precise) time of listening doesn't entail an equally precise analysis of the note. Waves can be modeled as sums of sines and cosines, per Fourier analysis. The variables at stake – the frequency at which a string vibrates and the time it takes to listen to this frequency – are “conjugate”, as they cannot be simultaneously measured (in one formalism, their operators do not commute). The more precisely we measure one of the variables – say, timing – the less precisely we can express its conjugate, frequency. The famous Uncertainty Principle expresses a mathematical inequality that pertains to the fundamental imprecision of complementary conjugate variables. This is just a mathematical fact about measuring Fourier duals. A better device wouldn’t do any good, since that reduces the measuring sample on one variable, at the expense of the measuring precision of the other – therein their complementarity.

What was remarkable for physicists a century ago was to translate that mathematical fact about measuring waves into a claim about reality. That’s gutsy metaphysics: to propose that reality, when involving complementary variables, presents itself in one of two incompatible ways: existing as one variable or the other, not both. Do processes in mind present complementarity? Scientific practice suggests that the way to go about answering that is seeing what it would buy us. Of course, even if one were to find situations of complementarity in mind, one need not exhibit the metaphysical guts that quantum physicists had, to thereby declare that mind is, hence, “dual”. That step in the reasoning is as drastic for mind as it has been for physical reality. Moreover, physically it is of course patent that brains propagate waves and their activities can and have been described in terms of vector fields. But that certainly doesn’t entail that the minds that emerge from those brains should then, ipso facto, be wave-dependent. Whether they are rests on how the mind-body problem is resolved, which may not happen any time soon.

### 3 Some Global Effects

A typical argument within quantum cognition works as follows. Suppose a subject can toss a coin to win $200 or lose $100. In Condition A of an experiment testing decisions made in
subsequent tosses, subjects are allowed to find out about the toss outcome before playing a second time; in Condition B, they are not. In both instances they’re given the option of tossing again. As Tversky and Shafir (1992) show for a version of this experiment, these results ensue:

(1) a. When believing they won the first round, a majority of subjects choose to play on the second.
   b. When believing they lost the first round, a majority of subjects choose to play on the second.
   c. When not given the results of the first round, the majority of subjects choose not to play again on the second.

Needless to say, (1c) is inconsistent with what happens in the majority of cases in (1a) and (1b). Quantum cognitive scientists account for this violation of the law of total probability as a quantum interference. Deviations from standard rational expectations can be understood by assuming that the overall conceptual landscape influences the subject's choice – the decision process is “context-sensitive”, in a sense to be discussed shortly.

Another example stems from understanding concepts as prototypes. A classic conundrum was posed in Osherson and Smith 1981: the typicality of a concept like a pet-fish (say, a guppy) is not a function of the typicality of the conjoined terms in the compound: a typical pet is experimentally ascertained to be warm and furry, while a typical fish is experimentally ascertained to be relatively large; the conjoined typicality of the terms is not predictable from the typicality of the concept ensuing from combining the two. Of course, this is only one of many problematic aspects compounds pose. So while it may be good to have a theory that predicts peculiarities in compound “composition”, it is important not to over-predict. Witness:

(2) I consider my fish a pet. My trout is definitely a pet-fish!

There doesn’t seem to be anything non-compositional here. Moreover, it would seem as if the fish and pet concepts in these sentences are dull exemplars. The point is: whatever is going on with compounds, one does want syntactic composition in the normal relations between arguments and predicates that sentential syntax instantiates.

With that as background, examine instances of so-called “amelioration” in syntactically unacceptable sentences. One such case was first raised by Haj Ross:

(3) They want to hire someone who speaks a Balkan language, but I don’t remember which (Balkan language) (?*they want to hire someone who speaks.)

While the example in full presents a violation of an “island” condition (extracting from inside a relative clause), the ellipsis (without the last parenthetical) is fine. A very different amelioration instance involves “that”-trace violations, as follows:

(4) Who do you think (*that) t died?

The trace of a Wh-extraction from subject position cannot follow an overt complementizer like that – though it is fine if the complementizer is dropped. Now witness:
(5) a. Who do you think that, **after years and years of cheating death**, t finally died? (Example by Barss and Deprez).
   b. Who did Leslie say that, to KIM, t had given the money? (Culicover)
   c. Who do you think that, **as we’re aware**, t loves silly books? (Ackema)
   d. Who does John doubt whether, **and Bill suspect that**, t cheated on the exam? (De Chene)
   e. Who do you think that t WROTE Barriers (as opposed to edited it)? (Drury)

None of their authors claim that these sentences are perfect, but they all indicate that the sharp violation in (4) ameliorates within them. The examples in (5) would seem to have the same structure as (4), with *that* as an overt complementizer and *who* as the extracted subject. In these instances there is “more discourse context”, although of various sorts: a “long” adverb in (5a), a displaced focalized phrase in (5b), an “as” interjection in (5c), right-node raising in (5d), contrastive focalization in (5e)… One final amelioration example involves the PRO-gate effect discovered by Jim Higginbotham:

(6) Who did PRO/*his kissing his mother upset?

The difference at stake here is rather trivial: pronunciation of the first *his*. When the pronoun is overt, though, the Wh-extraction leads to a Crossover effect; if the pronoun is dropped (replaced by PRO), the result improves.

Theorists have analyzed these situations in various ways. For example, Hornstein et al. 2007 accounts for the ellipsis amelioration by presenting a theory of islands that makes them sensitive to their phonetic shape. Hornstein and Kiguchi 2003 sees the PRO-gate amelioration as an argument that PRO may be the trace of A-movement, based on the fact that Crossover effects generally disappear with A-movement. Stories for the “that”-trace ameliorations have varied; the most recent is proposed in Sato and Dobashi 2013, which bases the account on the analysis of what counts as a prosodic phrase. These are three explanations for the three ameliorations.

Now recall the quantum interference approach to (1). Players that haven’t been informed of the output of their first coin toss can globally imagine both next steps, unlike if they have been informed of the fate of their first throw (which decides on heads or tails). In that global state, subjects are not considering a fate that ensues after they have “collapsed” the choice to either heads or tails; rather, both options are “live”, and therefore subjects make the “calculated” move not to play the second time around. Observe the possibilities:

(7) a. won $200; (i) won $200 = won $400 (.25 chances)
   (ii) lost $100 = won $100 (.25 chances)
   b. lost $100; (i) won $200 = won $100 (.25 chances)
   (ii) lost $100 = lost $200 (.25 chances)

Given these odds, one can debate whether taking a second toss pays off. But according to this approach, what gives subjects pause is *not knowing* what happened in the intermediate toss, which keeps all options for a further toss open, good or bad.

We may think of that state in terms of Coleridge’s famous “suspension of disbelief”, which keeps a subject engaged in a task for its duration, *regardless of its immediate plausibility*. With
that in mind, note that, for grammaticality judgments, a pattern holds across the ameliorations above. It can be abstractly represented as in (8):

\[
\begin{align*}
(8) & \ a. \ [\text{Wh-phrase} \ldots [\ldots t \ldots]\ldots] \\
 & \quad \uparrow \quad \uparrow \ ? \quad _R_ / \\
 & \quad b. \ [\text{Wh-phrase} \ldots [\text{PRO}]\ldots]\text{upset } t] \\
 & \quad \uparrow \uparrow \quad _R_/ \\
 & \quad \quad \quad ? \quad \quad \quad \\
 & \quad c. \ [\text{Wh-phrase} \ldots [\text{that} [X \ Y \ Z] [t\ldots\ldots]] \\
 & \quad \uparrow \quad \quad \quad \backslash \quad _R_\uparrow \\
 & \quad \quad \quad ? \quad \quad \quad \\
\end{align*}
\]

\(_R_\) in these examples denotes a context-sensitive relation of grammar that is independent of the one that yields the ungrammaticality. \(_R_\) in (8a) is ellipsis, it is control in (8b), and a variety of relations in the complementizer-trace effects, all complex. The question mark \(?\) in (8) denotes the offending relation, which ameliorates when its context of application overlaps with \(_R_\)’s context of application. This “suspension of ungrammaticality” is abstractly similar to the suspension of disbelief effect alluded to above, in that it is clearly a global process that “takes another look” at a bad example; so one may wonder whether this, also, could be seen as a quantum interference of some sort.

There are two distinct, though related, components to the possible approach being entertained. One is the very globality of the analysis. Of course, somewhat global conditions in grammar arise the moment one thinks of obviation effects, for example:

\[(9) \ \text{She feared (people thought…)} \ [\text{the FBI would kill Marilyn}].\]

Familiarly, she and Marilyn cannot co-refer, regardless of how far apart they are. However, there is a further twist to these ameliorations: the “suspension” effect appears to involve the computation of equally complex relations (\(_R_\) in these instances), which somehow puts things in a different light for speakers. Other ameliorations seem to take on that shape too, for instance “connectivity effects” discussed in Kayne (1984):

\[(10) \ a. \ ?*\text{What do you wonder why I bought?} \\
 & \quad \text{Who wondered why I bought what?} \\
 b. \ ?*\text{What did who see?} \\
 & \quad \text{What did who see where?}\]

The logical forms for the multiple questions in (10b) and (11b) are very similar to the ones involved in the simpler, albeit ungrammatical, examples (a Wh-island in (10a) and a Superiority effect in (11a)). Yet whatever causes the ungrammaticality in the relevant relation, as the new dependency \(_R_\) is introduced that overlaps with it, the results improve. It is, thus, the joint action of a certain globality in the computation, together with the overlap in context-sensitive relations, that yields the suspension of grammaticality. If there is interference here, in the quantum sense, it ought to involve both of those factors.
“Suspensions” (of disbelief or grammaticality) are strange, and the quantum cognition approach only gives us a way to talk about them. If the mind is acting in global terms in certain conditions, it is plausible to address these puzzles in quantum-like terms: by simultaneously considering the sum of representations being derived in parallel. Such an argument is strong only if alternative explanations aren’t and only as elegant as the predictive power of its underlying theory. To date there are no systematic treatments of amelioration, so this line of research is potentially interesting, and should be tested with examples constructed to fit the bill (ungrammatical context-sensitive dependencies that ameliorate in the context of “interfering”, equally complex, unrelated relations). Abstractly similar issues arise with grammatical preference, which typically involve situations as in (12):

(12) A boy saw every girl.
    a. $\exists x$, boy $x \left[ \forall y, \text{girl } y \left[ x \text{ saw } y \right] \right]$
    b. $\forall y$, girl $y \left[ \exists x, \text{boy } x \left[ x \text{ saw } y \right] \right]$

The structural ambiguity in (12) is well known. But despite the fact that both of those representations are grammatical, speakers feel that (12a) is preferred over (12b). To account for this, we need a comparison set and a procedure to establish the appropriate choice. If we find such a procedure within this framework, that would speak to its predictive power. In searching for this, we need to consider what sorts of grammatical conditions a quantum theory of syntactic categories and relations ought to involve.

4 Spooky Grammatical Paradigms

Consider whether familiar properties of grammars can, or should, be related to quantum considerations. Take simple projection, illustrated as follows:

(13) Chewing gum strengthens your jaws.

Of course Chewing gum can be a projection of chewing or a projection of gum instead. Projecting each lexical term means having one or the other determine the labeling of the ensuing combination. How does that process happen? Why is there not a “mixed” projection that has properties of both terms, instead? One way to actually predict this is by understanding the (chewing, gum) relation as involving, in some sense, conjugate variables, such that the verbal one is a dynamical variable tracking the spatial nominal one. Familiarly, one can think of a chewing as an event lasting while there is gum (left to chew), which can be modeled as the derivative over time of the mass term gum. If this is the right way to understand verb-theme relations, and this intuition is taken seriously in terms that have the dynamical (verbal) variable defined over the static (nominal) variable, it is plausible to treat the verb-noun relation as involving Fourier duals, with this consequence:

(14) Complementarity in head-complement relations
    For (head, complement) relations treated as Fourier duals, corresponding variables may be mathematically complementary.
This complementarity of complements entails that the complement variable can be specified, or the head variable – not both. That certainly tracks projection. For example, “collapsing” the relation in terms of the static (nominal) variable yields a nominal projection; concentrating on the dynamical (verbal) variable yields a verbal projection, which continues combining.

If (14) is the general approach to head-complement relations, the asymmetry of projection follows, as well as the fact that projections must settle on either a head-centered expression (head-complement relations, allowing further composition), or a non-head-centered expression (which terminates in that particular collapse, yielding predication). The formalism itself has as a consequence facts that otherwise need to be stipulated, e.g. as modes of semantic composition. The approach makes interesting predictions for other head-complement relations, for example predicting dynamical hypotactic (complementizer-centered) vs. static paratactic (clausede-centered) relations for complement clauses, or dynamical strong (determiner-centered) vs. static weak (noun centered) relations for determiner-noun relations.

Once complementarity is assumed for projection, similar questions can be posed for what is, in essence, a super-projection: the formation of movement chains. Can a chain also be considered a dual object, with a dynamical and a static expression, as it were? That is possible if all the steps of movement (the chain occurrences) are seen as superposed states of a dynamical object spanning over relevant syntactic contexts. The corresponding static variable is the one and only position in which such chain occurrences manifest themselves in interpretation, at PF and LF. A version of the Principle of Complementarity when applying to chains looks as follows:

(15) **Principle of Complementarity in Chains (PC\_C)**

Within derivation D, it is impossible to ascertain the scope \( \sigma \) of a given chain occurrence C while at the same time ascertaining the reach \( \rho \) of the superposed set of chain occurrences of C in D. Therefore, observation of derivation D with reach \( \rho \) collapses D to \( \rho \)'s conjugate variable \( \sigma \).

Where \( \rho \) is the set of contexts that the chain occurrences involve, \( \sigma \) is the corresponding scopes of those contexts. Assuming the variables are conjugate, the effect that the PC\_C has on PF representations is direct, where \( \sigma \) manifests itself as linearization. Consider in this regard Nunes’s 2004 analysis of the linearization of chain links, as in (16):

(16)  
   a. Gum was chewed t.
   b. **Gum** was chewed **gum**.
   c. Gum was chewed **chewed**.

If movement steps (copies boldfaced in (16b)) are identical occurrences, in chain-formation the upper and lower links of given chains are in the paradoxical situation of having to precede and follow intervening elements. Nunes concludes that the system resorts to “turning off” all copies but one, the linearizing element (as in (16c)). Suppose the set \{\{gum, T\}', \{gum, chewed\}\} (including the immediate syntactic contexts for each chain occurrence) is the derivational object with a certain reach \( \rho \). If \( \rho \) is the canonical conjugate of some \( \sigma \), observing \{\{gum, T\}', \{gum, chewed\}\} yields some value for \( \sigma \) in a PF linearization. This doesn’t tell us what the value for \( \sigma \) should be – e.g. gum preceding. But it is a significant start, as the conclusion that we need a definite value for \( \sigma \) follows from the PC\_C. In a nutshell: Interpreting syntax is observing it.
Identical considerations arise for the LF component, yielding what Hornstein (1995) calls conditions of “uniqueness” – the stipulation that multiple occurrences at LF are interpreted in a single location. This is more puzzling than it may seem at first, if the occurrences as copies are “real enough” to trigger reconstruction effects:

(17) a. Which pictures of myself [did John say t [Mary believes t [I took t]]]
b. Which pictures of herself [did John say t [Mary believes t [I took t]]]
c. Which pictures of himself [did John say t [Mary believes t [I took t]]]

The standard way to treat these sorts of effects is as in (18):

(18) a. [Which pictures of myself] [did John say [Which pictures of myself] [Mary believes [Which pictures of myself] [I took [Which pictures of myself]]]]
b. [Which pictures of herself] [did John say [Which pictures of herself] [Mary believes [Which pictures of herself] [I took [Which pictures of herself]]]]
c. [Which pictures of himself] [did John say [Which pictures of himself] [Mary believes [Which pictures of himself] [I took [Which pictures of himself]]]]

If the anaphor is local to a valid antecedent, different LFs arise, showing that, so long as the phrase which has matrix scope and the noun pictures establishes the variable position for the theme theta-role associated to took, where the restrictive material pictures of ...self is interpreted is up for grabs. That said, consider the ungrammatical (19):

(19) * Pictures of themselves seemed to no photographer to appear to no model to have been retouched.

Notice: a sentence like No man shows no woman bad pictures of themselves! demonstrates that anaphors like themselves can take split quantificational antecedents. In turn, Which picture of themselves would no man say no woman should destroy? shows that such split antecedents can be picked in successive-cyclic fashion. So then why can’t (19) be analyzed as in (20)?

(20) [[[pictures of themselves] seemed to no photographer
[[pictures of themselves] to appear to no model
[[pictures of themselves] to have been retouched ]]]

The PC C prevents the LF in (20), inasmuch as key to this object is the occurrence of themselves in two different intermediate sites. Uniqueness for LF occurrences impedes such an analysis – but it need not be stipulated: Uniqueness ensues as an LF consequence of a collapsed chain not being able to reach over two different scopes, as the PC C predicts.

The uniqueness of occurrences may seem to trivially follow from the fact that a single lexical element is involved, which “re-merges” at the appropriate (displaced) syntactic contexts. The problem with that approach is that it presupposes a notion of syntactic token that has no understood reality. Any syntactic theory must distinguish two identical token words within a sentence (e.g. the word word when we say this word precedes this other word) from various (syntactically arising) occurrences of these elements as above, and those two from the corresponding lexical type, say, word. The lexicon doesn’t contain each and every token of the
word *word* that has been used in this paragraph, but just a basic (concept, expression) pair that linguistically encodes that particular notion. The issue is how to activate that lexical type multiple times in a sentence, in some instances letting those activations survive to interpretation, while in others, whatever activation takes place within the derivation (as occurrences) dies out as the sentence collapses into an interpretation within the reach of that activation. The quantum approach takes the chain to be the sum of the probability of each occurrence – a copy at a given context – and has the uniqueness of tokens follow from the fact that interpreting one such dynamical object is interfering-with/observing it, which results in the collapse. A different token is a separate wave, which starts at the formation of the lexical array from the lexicon.

The consequences of the PC\textsubscript{C}, thus, seem reasonable for chains. They actually neatly extend to a treatment of the preferences in (12), if analyzed as in Martin and Uriagereka 2008. In the present approach chains are comprised of several *simultaneous* derivational stages: In their dynamic state $\rho$ they exist in the contexts the chain links span over. To interpret a chain with reach $\rho$ is to *collapse* the chain in scopal site $\sigma$. The chains in (12) can collapse as in (21) (where $\phi$ signals a Case/agreement site, $\theta$ a $\theta$-position):

$$
\text{(21)} \quad 
\begin{align*}
&\text{a. } [\text{IP a boy } [vP every girl } [vP \text{saw}]]) \\
&\phi \phi \\
&\text{b. } [\text{IP a boy } [vP \text{saw every girl}]] \\
&\phi \theta \\
&\text{c. } [\text{IP } [vP \text{ a boy } [vP \text{saw every girl}]]) \\
&\theta \theta \\
&\text{d. } [\text{IP } [vP \text{ every girl } [a \text{ boy } [vP \text{saw}]])] \\
&\phi \theta
\end{align*}
$$

75% of these representations collapse in an LF where *some boy* has scope over *every girl*; only 25% collapse in the inverse reading. These rates may correspond to individual speaker preferences. That makes sense only if individual speakers simultaneously access all the chain states in (21), which the quantum approach allows. An interesting question of course emerges in terms of what the nature of the preference set is (*where* “scope interactions” are possible). This probably ought to be related to the nature of the global domains where amelioration has been shown to obtain, but we set the topic aside now.

After having motivated the PC\textsubscript{C} and some of its consequences, we should ask also about notions that presuppose complementarity. Consider the following thought experiment. Imagine preparing two distinct syntactic categories-as-waves in such a way that they that start from “reduplicated” token waves within the same lexical array, of the sort arising in the emphatic (*these are*) pictures pictures (*not photocopies*). At some later time in the derivation, after movement, we have one of the prepared tokens collapse in a given site, while we have the other prepared, identical, token collapse in a crucially different site:

$$
\text{(22)} \quad 
\begin{align*}
&\text{a. PF: } \text{Pictures of themselves pictures of themselves seemed to no photographer to appear to no model to have been retouched.} \\
&\text{b. LF: } \\
&[[\text{pictures of themselves} ] [\text{pictures of themselves} ] \text{ seemed to no photographer to appear to no model to have been retouched}] \\
&[[\text{pictures of themselves} ] [\text{pictures of themselves} ] \text{ to appear to no model to have been retouched}]
\end{align*}
$$
Given the initial correlation between the prepared tokens (joined in the lexical array by their reduplicative identity) does this gambit not allow us to know both the reach and the scope in an object as in (22)? Of course that would violate the PCc, if meant ontologically. If we manage to determine the reach for a chain-wave C within a sentence, we cannot ascertain the scope of one of its occurrences, even in a roundabout fashion. This is good, as (22) is ungrammatical.

But now imagine the possibility that identical chains in these “prepared” reduplicative conditions were capable of instantly transmitting information, so that (surprisingly) the more definite the scope of one of the chain-waves is upon its collapsing, the reach of the other would then be more indefinite, and vice versa. While that may not yield a viable result for the collapses in (22), it may actually provide a rather interesting account for an old chestnut:

(23)

a. Lexical Array \{tried, \{to, leave, John, \{John\}\}\}

b. PF: John tried to leave.

c. LF: John tried [John to leave]

This analysis of control provides an alternative to Hornstein (1995)’s analysis. The key to the chain-like dependency is to start with a reduplicated pair (John, John) in the lexical array, only to use up one of the items in the first cycle, thereby to yield John to leave. Subsequently, in the next cycle, the second John in the array is used. This of course presupposes that a lexical array need not be exhausted in a given cycle: it may be possible to use whatever part of the array yields a grammatical representation, passing on the remaining materials up to the next cycle. This procedure yields the LF representation in (23c), which appropriately tracks the semantics in terms of thematic relations: there is a “trier” and there is a “leaver”. Of course, those two happen to be identical. But that aspect of the analysis is also captured if the two John’s are actually entangled, in such a way that, at the point of determining the chain’s reach, the system interprets the two John chains as a super-chain, with a unique derivational fate (signaled by the dotted line in (23)). Similarly, at the point of collapse the super-chain actually has a unique scope position, both at LF and at PF. While this super-chain does have chain characteristics, it does so because of entanglement, not because of complementarity (like regular chains).

Unlike complementarity, which operates on every local part of a construct, entanglement acts globally. So when we talk about entanglement on a sentence what we’re really discussing is an abstract, global, structure that is distributed throughout (a phrase-marker). In effect this is “action at a distance”, since the analysis does not commit us to the sorts of limitations (e.g., in terms of locality) that standard movement poses. Future research will need to determine the validity of such a consequence, as well as whether specific conditions limiting entanglement per se also obtain.

For example, the reason (22) cannot be salvaged as (23) is may be that in (22) there is no configurational theta-role to support the second pictures of themselves in a site separate from that of the first. In contrast, in (23) each John is in a valid theta configuration, as it enters the derivation. As a consequence of this analysis, one need not trivialize theta-roles as features (or otherwise giving up the good reasons why movement was traditionally considered different from control). While conditions of complementarity entail the standard chain behavior, conditions of entanglement that arise from complementarity allow for the chain-like analysis of instances that start as separate tokens in the lexical array. These notions (complementarity and entanglement) are part of the same machinery, but they are quite distinct. Complementarity is a property of waves
under Fourier analysis; entanglement follows from an ontology that depends on taking complementarity seriously. Entanglement presupposes complementarity, not vice-versa.

The present approach provides us, also, with a tool to analyze “partial” control, where the entanglement between the co-referential items is only partial:

(24) John asked to dance (together).

(25) a. Lexical Array {asked, {(together), dance, John, [ϕ], {[ϕ]}}} 
b. PF: John asked to dance (together) 
c. LF: [John [ϕ]] asked [ [ϕ] to dance (together)]

In this instance what is reduplicated is feature matrix [ϕ], essentially a pronominal clitic. While the first [ϕ] token in the array is used in the lowest cycle, satisfying the “dancer” theta-role, the second [ϕ] token, together with the category John, are passed on to the higher cycle – where they combine in clitic doubling fashion. Under these circumstances, the entanglement is between the clitic elements, so the co-reference has the weaker nature that pronominal dependencies exhibit more generally. Other chain-like phenomena – in terms of (PF, LF) interpretation, but not standard movement conditions – may be analyzed along the same lines – e.g. parasitic gaps, see Martin and Uriagereka 2014. Indeed, in principle any condition of identity of this “reduplicative” sort should be amenable to an entanglement approach. An intriguing such instance arises within ellipsis, which Chomsky 1995 speculates should involve “chain like” conditions. An entanglement approach may provide us with a way of illuminating the “parallel” effect in ellipsis, involving interpretive nuances that have never been explained.

5 Conclusions

This brief exploration of a syntax obeying complementarity is meant as promissory. The major theoretical move is bold, but scarce. Claiming that scope and reach within a syntactic chain are conjugate variables invites us to ask what in the nature of syntactic categories makes them be that way. It seems relevant that if movement (“internal merge”) lives on scope-reach pairs, in parallel fashion (regular) merge should live on the static-dynamic duality that manifests itself within head-complement relations. This provides support for why syntactic composition should be bottom up and why (firstly merged) complements should be central in articulating internal thematic and quantificational relations. The question is why the complex variables (projecting heads, reaching occurrences) are defined over more elementary variables (complements, scope). That parallelism doesn’t follow from anything understood, although it says something about how context-free computational relations constitute parts of context-sensitive ones.

Two notions of context are related here: the history of a derivation (or scope of a given occurrence) and context in the quantum sense that a dynamical system plays in determining the specific values of elementary duals. If syntactic context is physical context, one ought to ask whether other notions that are context-dependent, usually deployed in an information-theoretic sense, are another form of context that could be integrated. The question is then how syntactic context affects interpretive context, and “surface semantics” effects do suggest that a relation
exists. It would be good to examine, also, whether complementarity effects arise at that supra-sentential level, e.g. in terms of “discourse-linking” entailing amelioration of island conditions.

Computational elegance holds within sentence grammar, not below or above. Infra-sentential matters are known not to be systematic, transparent or productive. Supra-sentential matters have been formally studied, but it is unclear whether speakers have robust judgments about them, or discourse coherence becomes less reliable with size and speaker-specific conditions. Neither extreme is surprising for the systems explored here. Quantum cognitive theorists have explored the lack of compositionality within compounds or similar conditions about textual (in)coherence in some areas – which may be related to whatever happens in ameliorations – to argue for the quantum nature of the system. But just as things get strange below words or above sentences, finding quantum effects within sentence is clearly difficult, as sentences are well-behaved computational entities. It is only when going into reconstructions, construal, ameliorations and other such strange notions that a quantum system may be showing its face.

How pretty that face is depends on whether, in the process of assuming this architecture, one also wants to predict why the computation’s guts live in projected structures, displaced tokens, chain-like long-distance dependencies, and other such creatures. The debt owed to Noam Chomsky for having found such creatures and, more importantly, having provided a coherent analysis of their essential computational characteristics in Aspects, is only comparable to the most profound scientific breakthroughs of the 20th century, such as the quantum revolution.

Acknowledgments
This owes much to discussions with Massimo Piattelli-Palmarini, Doug Saddy, Ángel Gallego, Bill Idsardi and especially Roger Martin (who’s helped develop the ideas) and Michael Jarret (who’s helping to formalize them). Misunderstandings are the author’s responsibility. The debt to Noam Chomsky is impossible to express – one hopes these thoughts are not too unorthodox for his taste.

References
Chapter 3 of *Aspects* (especially pp. 131-137) presents two arguments for moving the “recursive property” of language from the transformational component to the base component. Although the arguments were compelling in the context of the findings in 1965, they are no longer, and in fact I think that something like the “generalized embedding transformations” of the pre-*Aspects* model may indeed be an essential part of an explanatory theory of syntax.

Prior to *Aspects*, the phrase structure rules of the base generated finitely many simple clauses; the “recursive property” arose from the operation of “generalized embedding transformations”, which embedded clauses from the base inside one another to build recursive structures\(^1\). So, for example, from \{ [John said S’], [he was sick]\}\(^2\) one could get \{[John said he was sick]\} by applying a generalized embedding transformation, which substitutes a clause from the base for the dummy marker S’. The diagram of embeddings in a derivation was called a “Transformation-marker”, by analogy with “Phrase-marker”. The other transformations, called “singulary transformations,” operated on the simple clauses of the base to derive their surface forms.

In the following passage, Chomsky observes that this arrangement permits derivations for which there is no apparent need:

> There are no known cases of ordering among generalized embedding transformations although such ordering is permitted by the theory of Transformation-markers. Furthermore, there are no really convincing cases of singulary transformations that must apply to a matrix sentence before a sentence transform is embedded in it, though this too is a possibility, according to the theory. (p. 133)

---

\(^1\) The fact that the base had NP/PP recursion was ignored in *Aspects* and will be ignored here also; but see Williams 1998 and Williams 2003 for considerations that suggest that clausal embedding is fundamentally different from NP embedding.

\(^2\) Braces enclose the “Workspace” in modern terms; in the pre-Aspects model, one might think of the base as the union of all Workspaces.
In order to eliminate these apparently unused possibilities of the theory, Chomsky proposes the model that we live with today:

These observations suggest a possible simplification of the theory of transformational grammar. Suppose that we eliminate the notions “generalized transformation” and “Transformation-marker” altogether. In the rewriting rules of the base (in fact, in its categorial component) the string #S# is introduced in the positions where in the illustrative example we introduced the symbol S'. That is, wherever a base Phrase-marker contains a position in which a sentence transform is to be introduced, we fill this position with the string #S#, which initiates derivations. We now allow the rules of the base to apply cyclically, preserving their linear order. (p. 135)

A unified base that encompassed all recursion, combined with cyclic application of the singulary transformational rules (p. 134), eliminates much of the indeterminacies of rule-interaction that existed before. Doing clausal embedding in the base fixes the interaction of clausal embedding and singulary transformations, and the transformational cycle fixes the order of singulary transformations applying in different clauses. Within a single clause, Chomsky observes that the singulary transformations might be partially ordered, whether by universal stipulation, or on a language-particular basis.

As a further argument against generalized embedding transformations, Chomsky suggests that the theory with a unified base makes semantic interpretation easier, as interpretation does not need to consult the Transformation-marker, as all relevant information about what is embedded in what is now represented in the base.

In order to regiment the ordering of singulary transformations, Williams 1974 proposed a “Supercyclic” theory. The cyclic theory suggested in Aspects says that no rule applies to an embedded clause once any rule has applied to the embedding clause. A Supercyclic theory says that no rule applies to an embedded phrase once any rule has applied to the phrase that embeds it. Supercyclicity is simply cyclicity with more granularity; in fact, with the most granularity. In the 1974 model there were four phrases that formed the “spine” of the clause: VP, PredP, S (= modern TP), and S' (= modern CP), and rules operating in different ones of these domains within the same clause were ordered with respect to each other; it remained, though, that rules operating within the same domain were not ordered with respect to each other by the theory. Supercyclicity thus removed some, but not all, of the looseness about ordering in the Aspects theory.

In the intervening years, the structure of the clause has enriched enormously. If we accept “cartographic” findings, instead of 4 layers of structure in a clause, there are over 100, and those ordered elements compose the “F-structure” of the clause. The result is that it is virtually guaranteed that Supercyclicity eliminates all of the play in the ordering of the singulary transformations in the Aspects model. If it looks like two different operations target the Spec of the same F_i, where F_i is an element of F-structure, then there is probably an F_{i+1} just above F_i that one of the rules targets instead. If this thinking is carried out thoroughly, there will be a one-to-one relation between F-structure and operations, and thus F-structure fully orders operations under Supercyclicity. With clausal recursion in the base and with Supercyclicity, all of the looseness in the ordering of operations is eliminated.

The story could stop there, but in fact Supercyclicity combined with rich F-structure offers a new opportunity to compare the two models of clausal embedding that are compared in Aspects. In Aspects, the choice was between “in the base” and “in the transformational component.” In
the modern context, where the base and the transformational component are not separated, the choice must be reformulated. A first attempt to resurrect the question might be put as a choice between these two:

(1) a. Clauses are Merged with the verbs they are complement to
b. One clause is embedded in another clause after both are formed

(a) is the “Base Theory” (BT) of clausal embedding in the modern context, the one that stems from the decision made in Aspects; (b) is the Modern Generalized Embedding Theory (MGET), the one that stems from the system that chapter 3 of Aspects puts aside.

A typical derivation of each is illustrated here:

(2) a. BT: \{ \text{think}_v, \text{[that Bill left]}_{\text{CP}} \} \rightarrow \{ \text{think} \ [\text{that Bill left}]_{\text{VP}} \}

b. MGET: \{ \text{[Mary \ [think]}_{\text{VP}}_{\text{CP}}, \text{[that Bill left]}_{\text{CP}} \} \rightarrow \{ \text{[Mary \ [think} \ [\text{that Bill left]}_{\text{CP}}_{\text{VP}}]_{\text{VP}} \}_{\text{CP}} \}

BT, like the Aspects model it descends from, is immune to the arguments in Aspects against generalized embedding transformations, since it doesn’t have them. MGET, on the other hand, raises again the questions about how the embedding operations are ordered with respect to each other and to other operations, and so on those grounds BT is to be preferred, and it is in fact the modern default standard.

However, the high granularity of F-structure offers an opportunity to adapt MGET in a way that nullifies the Aspects chapter 3 arguments. The richness of F-structure allows for a variety of clause types, where types are distinguished by “size” in F-structure terms. For example, suppose that 4 distinct points in F-Structure correspond to VP, vP, TP, and CP (in that order; that is, VP<vP<TP<CP). These define 4 sizes of clause, where, under the simplest assumptions, each is independently embeddable. vP might be the size of the complements of see, as in “I saw John move”; TP complements might be the size of complements to seem that allow raising to subject; and CP might be the size of full tensed indicative that-complements. Verbs will be specified for the size of their complement clauses: see for vP, seem for TP, think for CP. And given that F-structure is rich, with more than 100 points, there could be that many further clause sizes.

F-structure thus regiments three separate things: the structure of clauses, the order of operations, and the size of embeddable clauses. We are now in a position to answer Chomsky’s question about the theory MGET: how does clausal embedding interact with other operations? The more particular question to be answered is, when does clausal embedding occur? And since F-structure orders other operations, the pertinent question is, when, in F-structure terms, does clausal embedding take place?

The answer given in Williams 2003 was, “when the clause is big enough.” This answer makes sense in both BT and MGET. In BT it is trivial; if see takes a vP complement, we will build up the vP, and then merge it with see to make a VP:

(3) \{ \text{see, [Bill run]}_{\text{vP}} \} \rightarrow \{ \text{[see [Bill run]}_{\text{vP}}]_{\text{VP}} \}

---

3 For simplicity only complement clauses are considered here.
4 Braces enclose the “Workspace”; commas separate items in the Workspace.
5 In the modern theory with “internal” and “external” Merge, there is not the separation between “the base” and “singulary” transformations, so I have dropped those terms in favor of simply “operations.”
What is perhaps less obvious is that the answer also makes sense in MGET. In that theory, all clauses are generated as separate elements in the Workspace, and then are embedded in one another, as in the pre-Aspects model. The notion of “when” must be cashed out in terms of F-structure if we are to tie together the behavior of embedding operations and other operations, for, as we have determined already, the timing of other operations is fixed by F-structure and Supercyclicity. So the timing of clausal embedding has to be fixed by F-structure, but how?

As F-structure governs the structure of the clause, the size of embeddable clauses, and the timing of movements, it seems fitting that F-structure be thought of as more abstract than any one of these things. We might think of F-structure as a kind of clock that timed various operations. It is a clock which starts at F_0, and advances to F_1, then to F_2, and so on to F_n, and a derivation consists of a single sweep of the hand. Every action in the workspace is timed by this clock. As a consequence, all the clauses in the workspace “grow” simultaneously. In terms of the four F-structure points we have singled out, VP, vP, TP, and CP, this means that all clauses start out as VPss, then are built up simultaneously to vPs, then to TPs, and then to CPs.

With this assumption, it is perfectly clear what it means to say that a clause is embedded “when it is big enough.” If the embedding verb is *think*, which takes a CP, then the derivation goes like this:

<table>
<thead>
<tr>
<th>Workspace:</th>
<th>Clock:</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ Bill, think, he, lose } →</td>
<td>vP</td>
</tr>
<tr>
<td>{[Bill think]_vP, [he lose]_vP} →</td>
<td>TP</td>
</tr>
<tr>
<td>{[[Bill thinks]_vP]_TP, [[he lost]_vP]_TP} →</td>
<td>CP</td>
</tr>
<tr>
<td>{[[[Bill thinks]_vP]_TP]_CP, [that [[he lost]_vP]_TP]_CP} → embedding</td>
<td></td>
</tr>
<tr>
<td>{[[[Bill thinks [that[[he lost]_vP]_TP]_vP]_TP]_CP} → embedding</td>
<td></td>
</tr>
</tbody>
</table>

The embedding takes place at the very end, when the embedded clause is a CP. On the other hand, if the embedding predicate is *see*, which takes a vP-sized clause, then the derivation goes as follows:

<table>
<thead>
<tr>
<th>Workspace:</th>
<th>Clock:</th>
</tr>
</thead>
<tbody>
<tr>
<td>{ Bill, see, him, lose } →</td>
<td>vP</td>
</tr>
<tr>
<td>{[Bill see]_vP, [him lose]_vP} →</td>
<td>embedding</td>
</tr>
<tr>
<td>{Bill see [him lose]_vP]_vP} →</td>
<td>TP</td>
</tr>
<tr>
<td>{[[Bill saw [him lose]_vP]_TP} →</td>
<td>CP</td>
</tr>
<tr>
<td>{[[[Bill saw [him lose]_vP]_vP]_TP]_CP} →</td>
<td></td>
</tr>
</tbody>
</table>

As these derivations show, the embedding of the complement of *see* is “earlier” than the embedding of the complement of *think*. Williams 2003 called this timing of embedding in terms of F-structure “Level Embedding”; I will refer to MGET regimented by Level Embedding as simply “Level Embedding.”
fixed: two CP clauses will be embedded simultaneously, but a vP-sized clause will be embedded before a CP-sized clause. Second, the interaction of embedding with other operations is fixed as well, because now F-structure times all operations. The embedding of a clause of size F\_iP will follow all operations that target F\_i\_1 and smaller, and it will precede all operations that target F\_i\_1 and larger. So, all interactions are fixed by the theory, and Chomsky’s objection no longer holds. It is a fact that singular operations do apply to the embedding clause before the embedding takes place, and Chomsky claimed that “there were no really convincing cases” of that; but Level Embedding does nevertheless meet the spirit of Chomsky’s objection, because all ordering interactions are forced. And, as we will see below, the ordering of singularities to the embedding clause before the embedding has great explanatory potential after all.

So, Level Embedding is immune to Chomsky’s objections to generalized embedding operations, but is there any reason to prefer it to the scheme of derivation that we have called BT? Williams 2003 demonstrates that some general patterns of explanation are directly derivable from Level Embedding but not available in BT. These are first, a family of locality constraints, and second, a generalized “improper movement” constraint. In both cases, the explanations arise from the fact that under Level embedding, F-structure times everything—Merge, movement, clausal size, and clausal embedding.

To illustrate the kind of locality condition that follows from Level Embedding, we will consider subject raising of the sort that occurs with verbs like *seem*. It is unexplained in BT that subject raising is impossible from indirect questions, in that there seems to be no verb which takes an indirect question complement and allows subject raising out of it. In the following, *queem* is a made-up verb of the non-existing type:

\[(6) \ast \text{John} \_i \text{queems} [\text{whether} t \_i \text{to go}] \]

Level Embedding tells us that such predicates are impossible. It comes down to the ordering of two F-structure points: the point at which embedding takes place, and the point at which raising takes place. The point at which embedding takes place for questions must be the point that we have identified as CP, as questions always have a C; so, before that point, the structure of (6) is the workspace below, with the clauses still separate:

\[(7) \{ [[\text{it queems}]_\text{TP}, \text{[John to go]}_\text{TP}] \} \]

In order for the embedding to take place, CPs must be built up, since *queem* requires a CP:

\[(8) (7) \rightarrow (\text{CP step})
\]
\[{ [[\text{it queems}]_\text{TP}]_\text{CP}, \text{[whether[John to go]}_\text{TP}]_\text{CP} } \rightarrow (\text{embedding step})
\]
\[{ [[\text{it queems} \text{whether [John to go]}_\text{TP}]_\text{CP}]_\text{TP} } \]

The other relevant F-structure point is the point at which raising to subject takes place. Since raising targets Spec\_T, that point is the F-structure element we have identified as TP. This means that at the time at which raising must take place the sentence has the structure in (7); but raising cannot take place in (7), because the embedding has not occurred yet. Once the embedding has taken place, as in (8, last line), it is too late to do raising, as the derivation has passed on to point

\[^6\] The model also directly provides a generalized “reconstruction relation” not directly relevant here; see Williams 2003 for explanation.
The net result is that raising must always be from a TP, never a CP, and that is why indirect questions are always islands for raising.

We can easily generalize this kind of locality: if a movement targets F_i, then clauses larger than F_i will be islands for that movement. This follows immediately from Level Embedding; BT makes no such prediction.

In a similar way, Level Embedding rules out “improper movements” of the kind represented in (9):

\[
(9) \quad [\text{John} \quad \text{seems} \quad [\text{it} \quad \text{to go}]]_{\text{TP}}_{\text{CP}}_{\text{TP}}
\]

That is, a movement cannot relate the SpecC of a lower clause to the SpecT of an upper clause, even though the upper SpecT c-commands the lower SpecC.

This again follows immediately from Level Embedding: in order for the illicit movement to take place, the embedding must have already been done; however, that would mean that we have reached CP level operations, and it is then too late to perform an operation that targets specT. Again, we can easily generalize the conclusion: no movement from F_i to F_j is possible unless F_i < F_j, even if F_j is in a higher clause than F_i (and hence c-commands F_i). And again, BT makes no such prediction.

References

Does input data matter in language acquisition? Of course it does: a child born in Kansas learns English and a child born in Beijing learns Mandarin. But input data is hardly strings of words. According to the generative tradition of language acquisition outlined in *Aspects*,

> [T]he child approaches the data with the presumption that they are drawn from a language of a certain antecedently well-defined type, his problem being to determine which of the (humanly) possible languages is that of the community in which he is placed. (Chomsky, 1965:27)

That is, the primary linguistic data receives structural descriptions, which are determined by the strong generative capacity of the grammatical system under consideration (*Aspects*, §6 and §9).

While much attention has been devoted to the argument from the poverty of the stimulus (Chomsky, 1975), where the child demonstrates the mastery of linguistic knowledge in the absence of sufficient evidence (Legate and Yang, 2002), equally revealing observations can be made about the nature of the grammatical system when the child appears oblivious to an abundance of evidence in the input data. During the much studied null subject stage, for instance, English learning children systematically (and probabilistically) omit the use of the grammatical subject until around the age of 3 (Valian, 1991) even though adult speech almost always contains the subject. Children’s behavior would be puzzling if they interpreted the linguistic data as surface patterns, especially since we know that children are exceptionally good at matching the probabilistic aspects of language use (Roberts and Labov, 1995, Smith et al., 2009). If, on the other hand, children analyze the input under the guidance of “the (humanly) possible languages” (including pro-drop and topic drop grammars, Jaeggli and Safir 1989), a string such as *They play soccer* and indeed almost all the English data would not differentiate among these options (Hyams, 1986) and the prolonged stage of subject omission is accounted for (Yang, 2002).

In this paper we present a case study that highlights the input data’s explanatory limits when interpreted superficially, and explanatory power when given appropriate structural descriptions.
Our study concerns the acquisition of tense marking, and in particular past tense marking, in African American English (AAE) in contrast with Mainstream American English (MAE). Young children acquiring both varieties of English display a pattern, known as Root Infinitives (RI), whereby they produce a notable proportion of bare root verbs when a finite form is required. We show that the developmental patterns across dialects cannot be attributed to the statistical properties of specific linguistic forms but must make reference to an overarching and abstract system of grammar.

1 Input and Infinitives

The phenomenon of root infinitives (RI) has been and continues to be a major topic in language acquisition; see Phillips (1995) and Legate and Yang (2007) for review. For many languages in which the root verb needs to be tense marked, young children produce a significant number of verbs in the root clause that are nonfinite.

\begin{enumerate}
\item a. Papa have it. (English)
\item b. thee drinken. (Dutch)
\item c. Dormir petit bébé. (French)
\item d. mein Kakao hinstelln. (German)
\item e. Malon lauf. (Hebrew)
\end{enumerate}

An important aspect of the RI phenomenon is its gradient distribution, both within and across languages. Within a language, the use of RI is probabilistic rather than categorical, hence its alternative designation: “optional infinitive”. The nonfinite verbal forms are generally used in co-existence with finite forms, and children gradually, rather than suddenly, exit from the RI stage as they grow older. Furthermore, the cross-linguistic distribution of RI is also continuous: some languages have much longer RI stage than others: for instance, the RI stage may persist in English and Dutch for over four years while the stage is only very briefly present in languages such as Italian and Spanish.

The variational model of language acquisition (Yang 2002, 2004) can incorporate the statistical properties of the primary linguistic data within an abstract grammar model, and explains the gradient aspects of these properties. Instead of focusing on the deviation of child language from the adult form, the variational model interprets child errors as the attestation of non-target but UG-possible grammars or hypotheses (Pinker, 1984, Crain, 1991). Specifically, the learner associates probabilities with the grammatical hypotheses made available by UG. When a learner hears an input token, it probabilistically selects a grammar or a parameter value to analyze this token. If successful, the selected grammar is rewarded and its probability goes up; if the grammar fails, its probability is penalized or lowered. This type of learning scheme is very general, falling in the tradition of Reinforcement learning in psychology and computer science, and has been identified across domains and species.
In Legate and Yang 2007, we develop a variational learning account of RI. We attribute children’s nonfinite verbs to the type of UG-available grammar that does not mark tense on its root verbs, as attested for example in Mandarin. We call this the [-T] grammar, as contrasted with the [+T] grammar which does require tense marking on root verbs. If the English-learning child hears *I want water*, she cannot be sure that she is not learning a [-T] language, since the verb form of 1st person singular present is indistinguishable from the bare nonfinite form. On the other hand, if the English-learning child hears *I wanted water*, with overt morphological expression of past tense, she has reason to believe she is learning a [+T] language. Likewise, certain agreement verb forms (e.g., *John drinks coffee*), while not marking tense explicitly, is dependent on the presence of tense (secondary exponence; Carstairs 1987, Harley and Noyer 1999) thereby also providing evidence for the [+T] grammar. Given the abundance of you and me, and here and now in child directed speech, however, a significant proportion of root verbs that English-learning children hear do not show any evidence for tense marking. Hence, the child must rely on the morphological evidence for tense marking to learn that her language is [+T] and thereby unlearn the RI pattern for her language. Under the variational model, the speed with which the target grammar rises to dominance is determined by the frequency of disambiguating evidence that penalizes its competitor: given that languages offer different amounts of morphological evidence for tense marking, we predict that learners have different durations of the RI stage. Indeed, we found a significant inverse correlation between the amount of evidence for the [+T] grammar from tense-marked verbal inflections in the child-directed input data, and the duration of the RI stage across languages. Subsequent research has independently replicated our numerical results and, in some cases, found individual level correlation between the amount of tensed input and the rate of RI use between specific mother-child dyads (e.g. Rispoli et al. 2009).

It seems beyond doubt that the statistical properties of the input play an important role in the acquisition of language. It must be acknowledged that child language research in the generative tradition has rarely examined the quantitative aspects of linguistic input, which has given rise to the impression that input effects (such as the role of frequency) are inherently incompatible with the notion of UG and generative grammar. According to the alternative usage based approach, the child learner stores and retrieves lexically specific expressions in adult speech (Tomasello, 2000, 2003). Consider an alternative approach dubbed “MOSAIC” (e.g. Croker et al. 2000), which treats RI as the effect of purely input driven processing and learning. MOSAIC holds that the child stores lexically-specific combinations of words and other linguistic items, without the use of an abstract grammar. The child processes speech from right-to-left, retaining ever longer segments of the sentence: a sentence such as *She will jump* is retained as *Jump*, resulting in a root infinitive. Recent instantiations of this model (e.g. Freudenthal et al. 2010) have added an additional weaker left-to-right processing, in face of the obvious problem that early child speech does include initial wh-phrases and subjects, which would not be found through right-to-left processing. On a MOSAIC approach, there is no overarching abstract grammar and child language directly manifests the statistical composition of the input data. On the variational learning approach, the effects of the input are integrated into child language through abstract properties of the grammar. These differences lead to a stark contrast in predictions. For a variational learning model, verb forms that do not mark past tense, but rather mark agreement that is dependent on tense, do provide evidence for a [+T] grammar (as detailed above) and thus contribute to the child’s successful acquisition of [+T], for which past tense is a specific realization. In other words, hearing *She is jumping* or *She jumps* assists the child in her acquisition of *She jumped*. For a MOSAIC learning
model, in contrast, agreement that is dependent on tense cannot be relevant to the learning of past tense – hearing She is jumping would allow the child to store (She) jumping and She jumps would allow the child to store (She) jumps, but neither results in storage of (She) jumped.

Given this background, we turn in the following section to our test case: the acquisition of past tense in African American English (AAE) and Mainstream American English (MAE).

2 The Acquisition of Tense Across English Dialects

2.1 Tense in AAE

As is well known, AAE employs considerably less tense and agreement marking than MAE; see Green 2002 for review. The third person singular present -s is very frequently omitted, resulting in She jump. In addition, Labov’s (1969) classic study established the structural properties of copula deletion in AAE, as in She nice, including the important demonstration that copula omission in AAE obeys the same structural conditions as the contraction process in MAE She’s nice. Furthermore, AAE uses the bare form of be with a habitual aspect interpretation. All these verbal forms constitute evidence for a [-T] grammar and against the target [+T] grammar: AAE is superficially more like Mandarin than MAE. But these non-tense-marked forms are used in variation with tense marked forms; thus collectively AAE children sometimes receive evidence for a [+T] grammar, and sometimes receive evidence for a [-T] grammar. Compared with MAE, however, these properties of AAE reduce the amount of [+T] evidence available to the AAE learning child and increase the amount of [-T] evidence.

For simple past tense, however, the two dialects do not differ: AAE does mark past tense: She jumped. This fact is verified quantitatively in our analysis of child-directed AAE data presented below. Since AAE and MAE both mark past tense consistently, the grammar-based variational learning model and the usage-based MOSAIC make different predictions. On a MOSAIC model, the child carries out lexical learning by tracking and storing the usage patterns in adult speech. Therefore, we do not expect to find significant differences across dialect groups in child past tense marking, since the mothers use a comparable proportion of past tense tokens among their verbal forms. On a variational learning model, on the other hand, when the child learns tense, she is learning an overarching property of the language, [+/-T]. The MAE-learning children receive overall more evidence for [+T] marking than the AAE-learning: even though the past tense amounts are comparable, the MAE learner receives a good deal more evidence from third person singular, auxiliaries and copulas, which are frequently absent in AAE. We therefore predict that MAE children use past tense more consistently than AAE children, despite hearing a comparable amount of past tense data in the input.

2.2 The Acquisition Data

Our study is based on the Hall corpus (Hall et al. 1984) in the CHILDES database (MacWhinney 2000). We used the data from 35 children between the age of 4;6 to 5;0 and their mothers (four children from the Hall corpus provided very few data points in the recording sessions and were excluded). The Hall corpus consists of four demographic groups: black working (BW) class, black professional (BP) class, white working (WW) class and white professional (WP) class.
To determine children’s usage of past tense in obligatory contexts, we examined children’s utterances by hand. We judged the conversational context to determine whether the context requires the use of past tense, and then recorded whether the child had indeed used the past tense. This gives us the rate of past tense usage for every verb for the children. For the mothers’ data, we used the part of speech tagging and associated pattern-extraction tools used in the Legate and Yang 2007 study. The estimation of the amount of data in favor of the [+T] grammar from that study is consistent with results from an independent manual analysis of English data (Rispoli et al., 2009).

Table 1 gives a summary of the main results:

<table>
<thead>
<tr>
<th>group</th>
<th>child past %</th>
<th>past tense % in CDS</th>
<th>all tense % in CDS</th>
</tr>
</thead>
<tbody>
<tr>
<td>BW</td>
<td>87.2***</td>
<td>19.9</td>
<td>40.6***</td>
</tr>
<tr>
<td>BP</td>
<td>93.7</td>
<td>22.6</td>
<td>55.4</td>
</tr>
<tr>
<td>WW</td>
<td>94.4</td>
<td>17.1</td>
<td>49.8</td>
</tr>
<tr>
<td>WP</td>
<td>97.4</td>
<td>19.9</td>
<td>54.7</td>
</tr>
</tbody>
</table>

Table 1: Tense and past tense in child and mother’s language. The first column provides children’s percentage of past tense production in obligatory contexts. The second provides mothers’ percentage of past tense forms in all utterances that contain a matrix verb. Finally, the third column provides mothers’ percentage of forms providing evidence for a [+T] grammar in all utterances that contain a matrix verb (including past tense, and tense-dependent agreement).

Three notable findings emerge. First, there is no statistically significant difference in the percentage of utterances containing past tense forms over all matrix verb forms in the mothers across all four groups (column 2; $p = 0.89$). This confirms the observation that both AAE and MAE mark past tense consistently. Second, the BW class children produced a significantly lower rate of past tense than the other three groups of children ($p < 0.001$), which show no significant difference between them. Third, the BW class mothers produce a significantly smaller percentage of evidence for overall tense marking than the other three groups ($p < 0.001$); this reflects the well known differences between these dialects that we reviewed earlier. It also suggests that for the black professional class mothers, the dialect features of AAE with respect to tense marking are not very strong, and are in fact indistinguishable from the mothers in the white families in the data collected in the Hall corpus.

2.3 Input Mismatches and the Role of Grammar

Clearly, the differences in past tense marking between AAE working class children and the other three groups of children cannot be due to their mother’s use of past tense per se, since there is no difference between the mothers’ past tense usages. A series of correlation studies make the point more saliently.\(^1\)

Figure 1 plots the percentage of past tense forms in a mother’s speech against her child’s past tense usage rate in obligatory contexts.

\(^1\)For brevity, we only report the average rates of past tense marking for the 36 children. In the longer study under preparation, we provide results from mixed effects models with individual verbs and children as random effects. Our main conclusion is statistically confirmed: the rate of past tense marking in the children’s language is not determined by the rate of past tense marking in the mothers’ language, but rather by the overall evidence for tense marking in the mothers’ language, especially third person singular.
Notably, the past tense usage shows no correlation at all between the input and output as can be seen in the linear regression line in Figure 1. We conclude from this that the child cannot be performing lexical learning in past tense; the MOSAIC model, which stores (sub)strings from the input, cannot predict a difference between AAE working class children and the other three groups, because the past tense input data for all the children are not statistically different.

The overall tense marking evidence, by contrast, strongly predicts with children’s past tense usage ($r = 0.594, p < 0.001$). In fact, there is stronger correlation still between children’s past tense usage and the percentage of third person singular forms in the input, with a correlation coefficient of 0.68. Figure 2 illustrates the result of the linear regression.

Figure 1: Past tense marking in child and child directed language.

Figure 2: Past tense marking in child language and 3rd person singular marking in child directed language.
This demonstrates that the acquisition of past tense usage by children receives a boost from the overall amount of evidence for tense marking, even though the forms are not past tense per se. The effect of 3rd person singular in the input on children’s past tense is expected on the variational learning approach: since there is no difference in past tense marking across the input groups, third person singular is among the chief contributors to the cross dialect differences in the rate of overall tense marking, and hence the stronger predictor.

In the spirit of the discussion in Aspects, the child learner assigns a structural description to the input and the grammatical feature of [+T] is activated and reinforced whenever any tensed form is encountered. However, this connection between third person singular in the input and past tense production is entirely mysterious under the usage-based learning account, where the input is viewed as a string of lexical items. Under the grammar-based account, by contrast, they are connected, and work collectively, toward the learning of tense in the language.

3 Summary
We conclude with some general remarks on the acquisition of the tense and the role of the linguistic input in language acquisition.

First, in Legate and Yang 2007, we acknowledged that while analysts can count the input frequency of various forms, we had little understanding how the child learns these morphological forms, their constitutive parts, as well as their corresponding syntactic and semantic features and properties. We are in slightly better position now (see Lignos and Yang, 2010), but much remains unclear.

Second, our study reinforces the grammar-based approach to tense and RI, showing that tense is an overarching and systematic property of the grammar that is not acquired in a piecemeal fashion. The variational learning model, which makes use of very general probabilistic learning schemes, offers a general framework for evaluating the effect of input frequencies interacting with the grammatical model under assumption. When a correct model of the grammar is assumed, one which assigns the proper structural description to the linguistic data, the statistical correlations—as well as non-correlations—between child language and adult language follow. When a wrong model of grammar is assumed, or no model of grammar at all, they do not.

Finally, we hope that our project brings generative researchers’ attention to more quantitative studies of language variation and acquisition. Universal Grammar is compatible with input and frequency effects, which provide some of the best evidence for its validity. In addition to the inherent value in the study of dialects, important theoretical questions can be fruitfully studied and perhaps even resolved (Labov, 1969:761).

References


