Université de Montréal

Parsing Impoverished Syntax

Andrew Gordon Middleton

Département de Linguistique et Traduction

Faculté d'Arts et Science

Mémoire présenté à la Faculté des études supérieures en vue de l'obtention du grade de Maître és arts (M.A.) en linguistique

2006-08-31

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé:

Parsing Impoverished Syntax

présenté par:

.....

Andrew Gordon Middleton

a été évalué par un jury composé des personnes suivantes:

.....

Christine Tellier, président-rapporteur

.....

Daniel Valois, directeur de recherche

.....

Richard Patry, membre du jury

Résumé

Le langage servant surtout à la communication, un interlocuteur doit être en mesure de traiter la coréférence dans une séquence syntaxique donnée, faute de quoi il fait face à un problème de communication.

Alors que les marques de personne, nombre, et genre peuvent permettre de suivre la progression de la coréférence dans un discours, ce système est, du point de vue de l'interlocuteur, incomplet et irrégulier. Pour que l'interlocuteur puisse formuler la coréférence dans des phrases complexes, certaines conditions grammaticales doivent être remplies. Plutôt que de constituer des séquences soumises à certaines conditions grammaticales, les irrégularités linguistiques, telles que les îlots Qu-, peuvent être considérées comme des conditions qui bloquent la reconstruction et, par le fait même, la possibilité d'établir une relation de coréférence en bloquant le « mouvement » des syntagmes Qu- vers une position canonique.

Puisqu'il est possible de démontrer que les catégories lexicales (incluant les SN quantifiés) reposent sur des bases non arbitraires, ont peut remettre en doute le fait que, même si elles sont identifiables par le cerveau en tant que catégories grammaticales, elles fassent l'objet d'un encodage génétique (autrement dit, que la connaissance des ces catégories soit innée chez l'être humain).

Dans le cas de l'acquisition du langage par les enfants, ceci implique qu'ils ont peu d'indices visibles pour établir les relations d'arguments exigées par la composante lexicale, à moins que cette relation puisse s'établir sur des bases locales dans le discours.

En adoptant comme norme le traitement mental non séquentiel, cette localité des relations d'argument suppose une explication de l'acquisition des propositions adjointes différente de celle proposée par les théories faisant appel à la notion de Grammaire Universelle, qui stipule que ces propositions adjointes sont générées de façon indépendante des propositions principales et insérées subséquemment.

mots clés mouvement Qu-, indices, anti-reconstruction, localité, traitement non séquentiel, acquisition, propositions adjointes, portée des SN quantifiés, catégories lexicales, Grammaire Universelle.

Abstract

On the assumption that language is primarily used to communicate, if co-reference cannot be maintained within a given syntactic string by a listener, communication fails. While certain person, number, gender and case markings can be used to track co-reference in discourse, this system is—as manifested to a listener—incomplete and irregular. For the listener to reconstruct co-reference in complex sentences, certain conditions of the grammar must be respected. As such, apparent oddities of language, such as *wh*-islands, rather than being stipulations of the grammar, can be considered legibility conditions or anti-reconstruction effects. Thus *wh*-islands are points in the sentential string that impede 'movement' of *wh*-words or phrases to positions from where co-reference with a language-specific canonical position would be 'overwritten' by like elements, or otherwise impaired.

Lexical categories (including so-called quantifying NPs) can be shown to have non-arbitrary bases and so it is questionable whether they—even if identifiable discretely by the brain in terms of category—necessarily have an innate reflex there. The implication for child language acquisition is that there is little available to keep track of the argument relations resulting from lexical composition, unless there is a high degree of locality. Taking non-serial processing to be a regular brain function, this locality of argument relations implies an explanation of child acquisition of clausal adjuncts without the Universal Grammar hypothesis: adjunct clauses are generated independently of matrix clauses and subsequently interleaved.

keywords *wh*-movement, indices, anti-reconstruction, locality, non-serial processing, acquisition, adjuncts, quantifier scope, lexical categories, Universal Grammar

Contents

dedication I would like to dedicate this master's thesis to everyone I've ever annoyed by asking too many questions ... this is probably the fault of my parents

acknowledgements I would like to thank my advisor, Daniel Valois, for his important discussion, open mind and the freedom to develop some rather controversial ideas; I would also like to thank Kathryn Tippetts, Christine Tellier, Norvin Richards, Katarzyna Raczka and Kyle Johnson for their valuable discussion and encouragement; other people I want to thank, for sharing their time, insights, comments or otherwise making me think about these issues further, include: Jonathan Bobaljik, Cedric Boeckx, Željko Bošković, Danny Fox, Karina García Casanova, Brendan Gillon, Andrea Gualmini, Yosef Grodzinsky, Peter Hallman, Heidi Harley, Irene Heim, Hajime Hoji, Sabine Iatridou, Igor Mel'cuk, Raphael Mercado, Jon Nissenbaum, Ivan Sag, Yael Sharvit, Tim Stowell, Lisa Travis, Juan Uriagereka, Jean Roger Vergnaud, Maria Luisa Zubizarreta and Jan-Wouter Zwart. In the end and in all cases, all errors and biases are my own. I would also like to thank the people at the Département de Linguistique et Traduction at the Université de Montréal for their help, patience and financial support.

special thanks More than anything, I want to thank Catie and my family for their support. If not for them, where would any of this be?

Introduction

Noam Chomsky has occasionally used the story of an extraterrestrial, a 'Martian scientist', coming to Earth as a way of pointing out that such an outside observer would quickly note that one species has a communication system far in advance of all the others, and that this system of language patterns in non-obvious ways ((Chomsky, 1988, 41); (Hauser, Chomsky, and Fitch, 2002a, 1569)). To my knowledge, what Chomsky has not done is to turn the question around and ask what sort of language this intelligent extraterrestrial might use while making these observations. This seems an odd question if we are studying *languages*; we should be busy enough with attested languages and not need fictitious ones. If, on the other hand, we are studying *language* this question seems highly relevant. In a mind unconnected with our own genetic heritage could there exist something like our language?

Assuming this alien mind to be at least as sophisticated as our own but sharing no ancestry with us, what might we expect to find in this language? Does this mean that such a language would have perfectly regular morphology, lacking irregular verb forms, or perhaps lack such oddities as *wh*-islands or specified subject constraints? Morphological irregularities would not be so unexpected if this alien language has had contact with other alien languages and they had their own diachronic syntax. What might be unexpected is something similar to *wh*-islands. Nonetheless, this is only unexpected under current assumptions about Universal Grammar (UG) as a biological endowment particular to humans. Because UG has been formulated as a species specific means of explaining human language variation, the standard use of the term does not extend to explanations pertaining to the nature of information flow. Cases must exist when language is restricted because of such limitations. If, by Universal Grammar we meant 'the way language as an information system patterns regardless of the particulars of a biological endowment', we could extend it to any hypothetical language and perhaps better understand the subject of our inquiry. The nature of the biology of any particular species is no small factor in the ability to use recursive and sophisticated language. Nonetheless it is quite possibly not the critical factor in the form of such a language.

Specifically, we might ask whether this alien would have something like *wh*-islands or

even *wh*-words. Would this alien have our word classes? Would they have nouns or verbs? It would at the same time be surprising if they did but perhaps more surprising if they did not. Or rather, it would be expected that they could classify their environment into nouns and verbs and it might be expected that they could learn to query each other about this environment. Arguably, it might be unexpected is for these items to pattern in any way like our words and in particular like our *wh*-words and their island effects.

I continue with my story of an alien mind and ask that we consider the mind of the alien infant (henceforth AI) as analogous to a blank slate, in the sense that it has no specialized visual or linguistic component in its brain but simply an ability to note and recall patterns (e.g. in sound and in light) and track these over time. The AI has a great aptitude for classifying its environment but it has no 'built in' classifications. Suppose further that as part of the ability to classify things, the AI has a strong ability to see connections between the things classified (arguably necessary for the classification process itself). This ability is not unique to humans and many animals on our planet are able to understand signs that are only indirectly associated with their goals, such as a mating call and the mating act. In considering Pavlov's famous experiments with dogs, few would suggest that he somehow created a new biological function in the brains of these animals. He seems to have 'conditioned' a reflex that was already present in their neural physiology so that, specifically, the sound he chose became associated with the idea he chose. The physiological ability to disassociate sign from meaning—*signifiant* from *signifié* of Saussure (1916)—had to already be present in the dogs's brain.

According to these principles, in its earliest developmental stage, the AI classifies objects in its environment much as a human infant does. The AI has no prior knowledge of what should or could be labelled and has no notion of 'verb' or 'noun' and certainly no idea of what a 'quantifier' might be or how it might act. Nonetheless, it finds it useful to catalogue objects with *thing-referencing sounds* (henceforth *nouns*) and to organize the events or states these nouns are subject to with *event-referencing sounds* (henceforth *verbs*). These labels (henceforth *words*) seem to attain more subtle meaning (or 'implication') if placed in context with other words. It seems that verbs need nouns to have any concrete meaning just as cataloguing nouns is a pretty dull business without verbs to condition the meaning.

When the AI understands these word categories it can begin to understand the communication system in the adult language. The AI finds it useful to remember that even if the nouns have accompanying information conditioning or detailing its type (similar to our adjectives or relative clauses), they nonetheless pattern in relation to the verb. That is, even if the nouns do not directly precede or follow the verbs in a strict linear fashion, their occurrence is not random and, in deciphering parts of adult alien discourse, important to keep track of. Upon detecting a verb the AI awaits (or thinks back in time for) the occurrence of

nouns sufficient to make sense of just what was instrumental in the verb.

As it learns to process phrases of a simpler sort, the alien infant begins to see that there are other patterns of this nature. As it experiments with phrases of its own, it is essentially conservative, substituting one noun for another when it has perhaps heard them used in similar environments. Once it has made the connections between the words and their meanings, it does not connect other words into a phrase randomly. If utterances are about things and events affecting these things, there is no reason it would concatenate relational words (i.e. prepositions) or linking words (i.e. conjunctions) together. It has fought hard to understand the centrality of nouns and verbs; why would it experiment randomly with words that seem only to be used to add subtlety to these? There is no apparent reason a child would ever attempt sentences such as "beside at and the under dog" as some have implied. To presume that the child would experiment randomly if not guided by its biology is to blithely disregard how children proceed in any other learning task. Furthermore, scientific inquiry from the seventeenth century onwards has generally been rewarded by hypotheses assuming stepwise progression of events rather than portmanteau explanations in terms of particular or preternatural qualities of the subject at hand (Kuhn, 1962, 104). Arguing that the brain needs 'knowledge' (whatever might mean) of the almost occult quality of, for example, quantifiers, and that semantic meaning and syntactic form would not be possible without the guiding influence of UG seems something of a throwback to an earlier tradition of scientific deduction; for this very reason linguists urgently need to break with the traditions inherited from philosophers, logicians and traditional grammarians. Few modern scientists would leave their experiment unsupervised for something on the order of months or years and assume that environmental conditions had little or very minor effect on their experiment; yet this, in crude terms, is what psychologists and linguists do all the time.

Universal Grammar, as it is currently articulated, does not address the question of information patterning. As such, it is sometimes considered laughable to think that other entities with which we do not share a recent common ancestor (e.g. an alien or computer [1]) might use a language similar in general principles or sophistication to a human language. On the assumption that language conveys meaning (whether this is its primary purpose or not) and that meaning is another word for information, then linguistic science must consider patterns of information flow. As such, it is a disservice to our scientific ambitions to assume that an entity with sufficient powers of pattern recognition, and an ability to recombine these patterns, would necessarily produce gibberish, random or unrelated tokens or necessarily simplistic strings of meaning (etc.). This is to say, there is a sharp divide between those

[1] Representative of Strong Artificial Intelligence, i.e. a sentient and largely self-regulating intelligence known only in science fiction as opposed to makeshift and delineated projects such as the IBM computer that beat Gary Kasparov at chess

who do and do not believe that a computer might someday speak in a semblance of human language—without extensive doctoring—given some connectionist element such as pattern recognition (whatever that may turn out to be) and sufficient computing power. This connectionist element is likely something we have as part of our biological endowment but it also seems that other mammals as well as birds (and less sophisticated animals) have some sort of means of recognizing what is important to them and making some sense of the world. Quite possibly, intelligence is an emergent property of the ability to make a connection between any given x and y just as heat is an emergent property of the energy of molecules.

Too often the apparent 'answers' ascribed to UG have the flavour of a *deus ex machina*. If we allow that the nature of lexical categories are not in need of biological answer, but is one of several possible logical answers to the problem at hand (i.e. cataloguing and describing the world), is it possible that the way that they pattern in human languages are also possible or even expected outcomes of the same problem. Would the language of any species—given similar levels of intelligence and with similar articulatory and perceptual systems—exhibit any or all of the traits now associated with the Chomskyan Universal Grammar (UG) and Principles and Parameters (P&P)? If so, would this not suggest either that the traits associated with Chomsky's UG might be better attributed to *any* system of language?

Our scientists do not hesitate to assume that any advanced civilization will at some point develop mathematics. As such we have sent encoded messages containing examples of our mathematics into outer space in the hope that an alien civilization might intercept and understand the mathematics (and presumably know we have at least the beginning of civilization). If we assumed that mathematics were only possible in the context of the human brain—there are no other animals on Earth that can use anything but the crudest finger counting just as there are no animals that can use language except the crudest phrases—then it is almost impossible that any species without our biological lineage would have enough mathematics to interpret our codes [2] .

We agree that mathematics is not an arbitrary invention. Were we to lose our mathematical knowledge, we could, with time and perseverance, discover the laws again. Chomsky's UG hypothesis makes the assumption that language is for the most part arbitrary in its patterning. The hypothesis assumes in essence that, if there were a language spoken by an alien speaker, it would necessarily take a very different form than any human language (even assuming no difference in articulation system and working memory and intelligence). In these terms, UG by definition takes language to be largely arbitrary. This hypothesis seems sus-

[2] In fact, Chomsky in recent lectures at MIT has suggested that the language component in the brain might also be responsible for mathematics, a thesis that suggests that our Martian might also not have mathematics.

pect. This is not to pretend that all logically possible languages are to be found on this planet, but that the variation we might find in an alien speaker might be within the bounds of the variation found among human speakers. If language is largely non-arbitrary, what are the constraints we should expect language to be subject to?

How might we go about resolving these questions? If it is not to be a contest of opinions or received hunches, we must have a basis of agreement in order to proceed. To suppose that children do or do not acquire language particularly quickly is a matter of opinion. As such this claim is not resolvable and contributes little to the debate. That babies understand the continuity of objects or that children seem to understand ideas such as *inside*, *on top of* or *under* at too early an age to have 'learned' them is again a matter of opinion and cannot be taken as evidence of *conceptual primitives*—at least not without a basis of agreement. By the time we can test for these, children have been exposed to the physical evidence around them for an uncontrolled period of time.

Our ability to count is clearly, in some sense, a reflex of our biology, at least in the sense that our biology has given us the ability to distinguish breaks in continuous data and recognize objects as discrete entities. This is no small task, nor is the ability to count them. Nonetheless, Pepperberg and Gordon (2005) and Pepperberg et al. (1997) illustrate that these abilities are ones that we share with at least a few other species on the planet. Feynman (1961) demonstrates how, from the basic ability to recognize and count whole number integers, we can derive basic algebra. In similar fashion, we might ask, if from this ability to distinguish objects and count them we have developed our mathematics, can this means of deduction be applied successfully to word categories and ultimately the sometimes rather odd configurations of our syntax? If relational prepositions (*inside, on top of, under*) could not be understood without biological stipulation, then these are arbitrary. If these are arbitrary then we might need something like our biological heritage and UG to understand them because unlike mathematics their occurrence is more a product of adaptation to our biology than it is to general and observable principles of nature. That we buy shoes in pairs is arbitrary and not deducible to simpler principles than that we happen to have two feet. If these relational prepositions are arbitrary to me and I need specialized biology to understand them, then so should my dog and our Martian scientist. Assuming that I do not share significant or recent biological heritage with either of my dog or this scientist, and if we can suppose that it would not be beyond at least the scientist to understand whether it is *on* or *off* the couch, then perhaps this is non-arbitrary. The necessary accidents of evolutionary biology to provide at least these three species with such knowledge is unforgiving to an innatist hypothesis.

The primary question under investigation in this essay is whether the known configurations of syntax have possible non-arbitrary explanations. We might consider that the

non-obvious patterning of syntax as a necessary result of how information patterns, for example, in the absence of a rich system of co-referential indices. We might then ask if this non-obvious patterning is in fact in need of biological stipulation. If not, if they can be deduced from the logic of the problem of how to communicate moderately complex meaning with spoken sound over time (or in the case of sign language, signs over time), then perhaps we might reconsider our approach to the scientific investigation of language. It is not a question of finding the realm of the possible, what our genes might allow and what analogous complexity one can find in nature. It is a question of accounting for the data with the least unlikely explanation.

Chapter 1

Neurological Considerations & the Principle of Learning

> *Scientists sometimes deceive themselves into thinking that philosophical ideas are only, at best, decorations or parasitic commentaries on the hard, objective triumphs of science, and that they themselves are immune to the confusions that philosophers devote their lives to dissolving. But there is no such thing as philosophy-free science; there is only science whose philosophical baggage is taken on board without examination.* Dennett (1995)

1.1. The Nature of the Problem & the Shape of the Debate

Recent investigations of syntax in Chomskyan linguistics, such as Chomsky (1995b), have become very involved in the question of the so-called *architecture of language* as it might pertain to issues of human cognition. That is, the human ability for recursive language is seen as a discrete system and one that might be representative of other such discrete systems of cognition in the human brain. Chomksy's hypothesis of some kind of a discrete 'language organ' responsible for such a system has been controversial since it was proposed in Chomsky (1965). In this essay I argue that both Chomsky and his critics have become distracted by questions of what language is about and how it gets that way. Chomsky's recent writings seem more concerned with engineering his proposed system into as elegant a machinery as possible, often focusing on a hypothetical level of language called Logical Form (LF) while others say, *But that isn't how language works...* perhaps preferring to think that the *form* of language as something due to intention, motivated by semantics or formed by culture.

Specifically, while I argue against the notion that language is a discrete computational system, so too do I argue that Chomskyans and their critics continue to hold to many seemingly metaphysical notions about word meaning and syntactic categories (etc.). By this I mean that, while a given word, such as the quantifier 'any' or the *wh*-word 'what' or even

the common noun 'dog', might have very complex meaning on analysis, this is not to say that it necessarily has this rich meaning in any discrete sense when used by the brain in language. While a given verb might be associated with certain nouns or be able to take an indirect object, it does not follow that there exists a generalized 'argument structure' (Hale and Keyser (2002)) discretely archived in the brain. Nor does it follow that, once able to use *wh*-words (what, why, where etc.) or generalized quantifiers (some, any, most etc.), the brain is liable to perform strange operations such as *wh*-movement (Chomsky (1977), Richards (2001) and references within) and *quantifier raising* (QR) (May, 1977).

As an example of what I refer to as metaphysical properties of language is how we analyse or describe the composition of verbs with their nominal arguments. Verbs are sometimes defined as a lexical category in need of saturation by nouns (Baker (2003)). Nonetheless, this presumes that verbs are a preternatural category and that this category is recognized by the brain and in turn that this category is determined by its need of saturation by whatever nouns might be. This is to ignore the fact that, by the nature of how verbs are used in language, they have no recognized means of referring and are consequently largely meaningless without elements that do refer, i.e. nouns. That nouns themselves are used to refer follows from basic principles such as the *arbitrariness of the sign* (Saussure (1916)). This is itself not a uniquely human ability, given that at the very least other mammals are able to make associations between arbitrarily related *signifiant* and *signifié*, for example the dogs in Pavlov's famous experiments. While the observation that verbs can be defined as lexical categories in need of combination with nouns (so that they might refer) may seem very much like the idea that they need saturation by nouns, it is critical to distinguish between descriptions [determined by how language is used] and those that assume innate and seemingly arbitrary and discrete properties of lexical categories.

The basis of this essay, which relates, and argues against, discrete forms and functions of grammar with more so-called metaphysical characteristics of grammar (*wh*-movement, feature checking, QR etc.), is the *impoverishment of syntax hypothesis*. This essay presumes that natural human language syntax is fundamentally *impoverished* on at least one but more likely two significant levels, the level of spoken (or signed) language and the level of mental processing and storage (or representation). Of these two, the impoverishment of spoken syntax is the focus of this essay because, I argue, it is the level that most directly affects the forms taken by syntax and for which there is clear empirical evidence. For the second level of impoverishment, the level of mental processing and storage, there is less direct but nonetheless compelling evidence discussed in such works as Mountcastle (1982).

The impoverishment of syntax at the spoken level refers to the fact that, because spoken indices and word categories are poorly indicated at best, listener comprehension—or listener failure of comprehension due to the absence of these—necessarily conditions the

forms taken by syntax. This is to say, while certain elements, such as 'phi (ϕ) features' (person, gender and number morphology) (Bejar (2003) and references within) and morphological case, might provide a means of relating arguments to verbs (especially when displaced), these are often insufficient or at least too inconsistent for a listener to reliably parse the syntax if effects such as locality (island effects, so-called barriers or phases etc.) did not also condition the grammar. Furthermore, because there is no consistently distinguishable means of recognizing lexical categories (verb, noun, *wh*-word, quantifier etc.) as part of the audible signal, it is doubtful that any hypothetical function of the brain would be able to take advantage of processing by category. This is to say, if there is no particular audible *signature* or *identifier* marking membership in a lexical category, it is doubtful that such functions as feature checking of verbs, *wh*-movement, QR etc. could function in the context of a discrete system of grammar. As such, this essay assumes the forms of language to be determined largely by non-arbitrary by-products of the limitations of interpretability caused by the impoverishment of spoken language.

These problems of indices and identifiers are not recognized or, at least, not taken to be fundamental to language forms by Chomskyan theory. Instead the theory largely considers the forms of grammar to be determined by a discrete system of computation by a discrete language processing area of the brain. The forms of language are taken to be arbitrary and without explanation other than by such a discrete system which is presumed to be determined by the biology.

Nonetheless, Chomsky's original question is critical to any serious investigation of syntax. By this I mean the question: *how does syntax attain the form that it has?* That is, the particular forms manifested by human syntax may not be important in terms of cognition and may never tell us very much about the workings of the human brain but this is not to say that the forms taken by syntax are not worthwhile subjects of scientific investigation. Nonetheless, the question of how or why syntax takes the manifest forms it does has become something of a secondary issue. In effect, the proposed answers have taken over the debate and the questions that originally prompted them have been all but forgotten.

Why can we not say (1a) unless we mean to inquire into how one might wonder about Catie degreasing her bike chain. Why can it not mean (1b)?

(1) a. * How do you wonder why Catie degreased her bike chain?

b. Why do you wonder how Catie degreased her bike chain?

This question seems tacit in much of Chomskyan linguistics but I believe it has been replaced by the question: *what function of the discrete computational system (known as Universal Grammar (UG)), that is innate and biologically determined, is responsible for being able to say (1b) and not (1a)?* I take this to be a very different question. In fact,

much of contemporary linguistic theory operates on the assumption that in order for natural human language to exist in the forms that it does, there necessarily needs to be some part of the brain that is innately specialized to account for human linguistic competence. While this assumption is controversial, it is generally considered to be the only tenable working assumption within the area of Chomskian linguistics. This essay challenges the merits of this assumption[1]. Chomsky (1965) proposed that the human capacity for language is most likely due to some kind of 'language organ' roughly analogous to organs in the body. This theory has been very popular in a number of academic fields despite the fact that there seems to be little evidence of the necessary physiological specialization in the neo-cortex (Mountcastle (1982)). This is not to assume any particular specialization but simply that if this theory is to have any meaning there must be important differences between, for example, the cortical areas responsible for vision as compared with those responsible for language; if there are not such differences then we need to reassess what one means by a theory of innateness.

As Chomsky himself frames in in *The Architecture of Language* (2000), the fact that we can learn language is to say that that language is necessarily innate to humans:

> There is a huge literature arguing against the innateness of language; there's nothing defending the thesis. So the debate is kind of funny in that it is one-sided. Lots of people reject the proposal that language is innate but nobody ever answers them. The reason why nobody answers is that the arguments make no sense. There's no way to answer them. To say that language is not innate is to say that there is no difference between my granddaughter, a rock and a rabbit. In other words, if you take a rock, a rabbit and my granddaughter and put them in a community where people are talking English, they'll all learn English. If people believe that, then they believe that language is not innate. If they believe that there is a difference between my granddaughter, a rabbit and a rock, then they believe that language is innate. So people who are proposing that there is something debatable about the assumption that language is innate are just confused. So deeply confused that there is no way of answering their arguments. There is no doubt that language is an innate faculty.

One trouble with this line of reasoning is that it undermines the basis of the discussion; in order to advance our understanding of the question, we need to evaluate the question without resort to hyperbole. An interesting line of inquiry might be what distinguishes african grey parrots[2], chimpanzees and Chomsky's granddaughter. Here we can begin a

[1] And, as a secondary hypothesis, the general tendency to expect there to be more specialization in the mind than not

[2] See, for example, work by Irene Pepperberg of Brandeis University.

scientific inquiry into the differences and perhaps learn something about our facilities for language. In a fair and objective inquiry, we are best to avoid reducing whatever the answer may be to an obvious result[3]. What is does not do is answer the question that is of interest: the question of how does syntax attain the form it does and how might the human brain process language and other information.

If we are interested in this question, we are necessarily interested in the details of the answer, not in whether one personality or another is more persuasive in their general description. Therefore the first part of this essay will necessarily describe what the problems with this hypothesis seem to be and the second part will address certain traditional data problems that have been held up as 'if not innate then what?'

1.1.1 Constraining the debate

The debate needs to be constrained by a number of factors. An important contribution of Chomsky's has been to focus talk about the 'mind' as being a function of the brain and a concern for someday understanding functions of the brain as we seek to understand functions of the body. We do not—and will not for some time—know, for example, how protein signalling within neural cells work together to ultimately allow us to understand the Gettysburg Address. That is, because we cannot for some time know how it is that the brain processes linguistic or visual information, we should be careful about proposing 'null operators' or strange attractors such as Case. For that matter, as controversial as it sounds, we cannot necessarily assume that language is fundamentally a logical, rule-based, set of operations[4]. The apparent need for such things as 'null operators' could be interpreted as a red flag pointing to the fact that language is not rule based; that is, if we cannot reconcile the structure of language without positing such things as null operators, perhaps we are on the wrong track[5]. In light of these limitations and as linguists, we need to try our utmost to see what functions of language can be accounted for in terms of information processing. In a spoken stream of language, what information will effectively be 'overwritten' in the soundstream and so unavailable for listener interpretation. This is to say, what information—or lack of—in the manifest syntax can account for the 'non-obvious' patterning of syntax?

[3] Regardless of whether the 'language organ' hypothesis is shown to be accurate, language is still innate to humans—here there is a subtle shift of course from whether language is innate to the human brain, i.e. part of our genetic inheritance, or a combination of an aptitude for something like language and other factors including historic accident (when it pertains to arbitrary elements often referred to as 'parameters') but more importantly the legibility of the linguistic string

[4] While in Chomsky (1981) (§4.1 "The Variety of Rule Systems" etc.) there is an effort to distinguish between 'rules' and 'principles' the distinction is more one of specificity than kind; the problems I list as regards rules pertains equally to principles in the sense generally used in Chomskyan theory; i.e. in that the principles are stipulations of the grammar and seem to function in an 'if not x then y manner' principles can be considered as rules in this sense.

[5] Furthermore, making too strict analogy to computers or rule of law is of course problematic. If our minds are computer-like, where is the code? Perhaps language is more to do with 'pattern regulation' than rules per se.

Only once any influence from these factors can be accounted for, can we feel justified in proposing that these patterns are caused by genetically caused innate constraints. Linguistics constraints may in fact be innate to language more than they are innate to humans.

1.2. Logic or science?

In 1955, Chomsky proposed in *The Logical Structure of Linguistic Theory* (LSLT) that natural language is governed by rules that should be decipherable by using the tools of the logician. That is to say, human language can be adequately described by determining a set of logical rules that could generate all licit utterances of a language while excluding all illicit utterances. As argued in Chomsky (1965), "Clearly, a child who has learned a language has developed an internal representation of a system of rules that determine how sentences are to be formed, used and understood" (25).

Although Chomsky added the insight that human language is a product of the brain and, as such, human language must be studied within the context of a biological system, it is important to consider whether this approach to the study of language should be considered a *continuation* or a *break* from more traditional approaches. That is to say, while Chomskyans[6] can be described as taking the rule-based analysis of language to its logical conclusion by trying to determine what biological systems might provide the constraints and operational mechanism of language, it does not follow that this is the only or best way to study language as an empirical problem. Most of the debate by linguists or psychologists tacitly assume that there is some significant level of rule-based operations that must be in place innately or learned by a speaker. The belief that language is largely compositional is common also to traditional grammars and lay conceptions of language. Language is thought by many if not all investigators to be in some way or another the computation of discrete elements on the basis of a system of rules, whatever those may turn out to be. Chomskyans have maintained that this is only possible if there is a biologically unique area (or areas) in the human brain that is specialized for human language and that words (or morphemes or phrases) are stored in something analogous to a *lexicon* and in some way this feeds the syntactic computation. Both form (syntax) and meaning (semantics) are the

[6]By which I mean the variety of schools of linguistics that could accurately be called *generative*, and that ultimately derive from Chomsky's early proposals, including but not limited to Principles and Parameters transformational grammar (i.e. Government & Binding; Minimalism), Optimality Theoretic grammar (OT), Head-driven Phrase Structure Grammar (HPSG), Lexical Functional Grammar (LFG), Categorial Grammar (and combinatory categorial grammar (CCG)). It should furthermore be noted that this seeming embarrassment of riches is a serious problem for the basic theoretical assumptions pertaining to all of the versions of the theory. As remarked in Kuhn (1962) "proliferation of versions of a theory is a very usual symptom of crisis. In his preface, Copernicus complained about it as well" (71); see also Kuhn (1957). The problem is also mentioned in Greene:2001, The Elegant Universeregarding the crisis in String Theory; if there is no more or less objective way (empirically or theoretically) of choosing one from another, basis of distinguishing between theories comes down to individual scientific 'choice' and educational background.

assumed consequence of this process[7].

In this essay I work on a different hypothesis. It is entirely possible that Chomsky and other generativists maintained the existing view that language is essentially a rule based system without considering that, as a feature of human cognition, it may only superficially, or in part, adhere to rules of various kinds but have at its root another basis of operation. The rules that Generativists have apparently found are important for linguistic sense, I attribute to *the nature of the problem*[8]. These 'rules', pertaining to parasitic gaps (PGs), *wh*-word configurations etc., may be the *emergent properties* that would exist in any linguistic system if it is used by any creature limited to spoken and heard language exchanges[9]. The physiological foundation of this basic ability, I suggest, is likely to be based in the functions of the neocortex. My starting assumption is directly at odds with that of Chomsky because it assumes a central role to pattern recognition and some physiologically based analogy (as seems available in general cognition). As pointed out in Lewis and Elman (2004) "Statistical learning, and 'any account which assigns a fundamental role to segmentation, categorization, analogy, and generalization' is rejected in Chomskyan linguistics as 'mistaken in principle' (Chomsky (1975))". I argue that the opposite is true. More precisely, I argue that the biological foundations of language are quite probably exactly these (and must ultimately be invested with the tools of the biological sciences); the 'non-obvious' effects such as *wh*-movement that Chomskyans consider more pressing, may very well be the result of non-biological and secondary effects (see §5.1). That is, a language could conceivably exist without *wh*-movement (and Chinese is an example of this) but it is unlikely that a language could exist without segmentation, categorization, analogy, and generalization (see §2).

1.2.1 The Mountcastle Hypothesis

In Mountcastle (1982), the eminent Johns Hopkins University neuro-physiologist Vernon B. Mountcastle observes that, contrary to contemporary assumptions, *the neocortex shows little evidence of specialization from one region to the next.* This is to say that the details neuro-physiologists have found from region to region are not significantly different to account for the differences attributed to them. From this observation he proposes that the

[7]The details of whether or not this is one contiguous 'organ' is quite beside the point but the question whether there are some regions in the brain that react specifically and uniquely to language and that these areas behave in this way due to some kind of specialization and not due to the fact that they are possible but not unique candidates for such a role is the question I want to focus on.

[8]I assume that there are other phenomena such as the definiteness effect that may in fact are *perceived* as rules or conditions of language because of our tendency to analyse the whole as intentional and rule based. These elements may be the result of accidental correspondences that came to signify subtle differences without ever having more than an associative significance.

[9]The term *emergent property* is defined in the *Oxford American Dictionary* as "**1** PHILOSOPHY (of a property) arising as an effect of complex causes and not analyzable simply as the sum of their effects ... **2** arising and existing only as a phenomenon of independent parts working together, and not predictable on the basis of their properties: *one such emergent property is the ability, already described, of an established ecosystem to repel an invading species.*"

neocortex must have a common function whereby essentially the same cortical structures are responsible for the various known workings of the brain. Under this hypothesis, the neural tissue pertaining to visual data in the visual cortex is essentially the same and performs essentially the same function as areas associated with language etc. That is, there is a general function of the cortex that accounts for the manifest workings of the brain without the various regions being specialized for the tasks associated with them. The visual cortex only stores and reacts to visual data because it is connected to the optic nerve (etc.) not because there is anything essentially *visual* about the cortex in that region.

1.2.2 'Strange' Conditions

This is not to deny that there are conditions that language seems to adhere to, such as conditions on parasitic gaps (PGs) or *wh*-word order and configuration. It is simply to question whether these conditions on configurations of the syntax are stipulated by the biology or whether they are due to other causes. Furthermore, it is vitally important for even the most abstract theorist to correlate the theory with what (albeit limited) is known about the tissues in the neocortex that might account for the phenomenon. Are conditions on PGs or 'subjacency' emergent properties due to the nature of the problem or are they deeper properties of language?

In discussing the pro-drop parameter in *Lectures on Government and Binding* (LGB), Chomsky introduces the topic in the following way: "The most interesting topic in connection with the RES(NIC) is the clustering of properties related to the pro-drop parameter, *whatever this turns out to be*" (240, emphasis added). This hedge is noteworthy both because it suggests that at the same time as advocating a theory of Parameters, Chomsky understands that the phenomena under discussion may well later be understood in other terms, terms that may in fact show these, while not isolated phenomena, are phenomena that may all be effects of factors *other* than that of a single parametric switch triggered in the child's mind.

This is not an isolated comment. Earlier in LGB, Chomsky makes this belief clear: "To mention another case, I will suggest that the *[*that*-trace] filter of Chomsky and Lasnik (1977) is too 'strange' to be an appropriate candidate for UG and should be reduced to other more natural and more general principles.... Similarly, I will suggest that the two binding principles of the OB-system—the (Specified) Subject Condition SSC and the Nominative Island Condition NIC—are implausible because of their form, and should be reduced to more reasonable principles" (14).

This of course is what most of LGB was about and why it had such a great impact on modern Generative linguistics. These phenomena were "too strange" to be hardwired into any possible language organ. In LGB, Chomsky attempts to justify many of these apparent

effects by means of the principles of Government and Binding. Government proved later to be such a troublesome thing to define that it was famously abandoned in the Minimalist Program (MP).

Elsewhere in LGB (Chomsky, 1980, 67) Chomsky discusses the idea that parameters show variation in languages much like Jacob (1976)) discussed for the natural world of animal variation ((Boeckx and Piattelli-Palmarini, in press, 8)). The idea that there might be parameters that might work as biological switches in the context of Universal Grammar (UG) in conjunction with principles (or 'rules') became a small revolution within transformational theory. These parameters are often thought of as biological switches triggered by exposure to particular linguistic data.

Nonetheless, another way to look at parameters is as *descriptive generalizations*[10], where the details of a particular language (rich morphology e.g. Italian or conservative argument configurations e.g. Chinese) allow for such phenomena as *pro*-drop.

1.2.3 The (Relevant) Logic of the Problem

While there has been much effort to understand how human language might be a logical, rule based system and concern for a so-called Logical Form and the stipulative and questionable notion of 'logical argument position', there has not been comparable effort to examine the logic of the problem. Upon recognition of the biological nature of the problem there is, to my knowledge, little effort to understand how the physiology of the neocortex might work in tandem with secondary, non-neural, factors to account for the linguistic data (i.e. constraints arising due to the nature of the medium of primarily oral communication or its analogues). Instead it was proposed that one system, presumably biological, must account for *all* the details of the data at once. To better understand the role and character of our biological endowment (and how this pertains to language) might be simplified if we separate the problem in two parts.

(2) *internal elements (i.e. physiology)*: cortical tissue capable of perceiving, repeating and permuting patterns; articulatory system; perceptual system

(3) *external elements (i.e. what is perhaps inherent to* any *system of language*: system that must conform to certain limitations of the physiology involved; due to various limitations of this physiology and the need for some degree of utility, this system chooses to parse strings in shorter chunks that are only crudely indexed (case, gender, number etc.); more sophisticated indices would in theory be possible but this option seems not to be taken, possibly in favour of some efficiency associated with relatively smaller packages (i.e. sentences, perhaps with subordinate clauses)

[10]Thanks to Norvin Richards (personal communication) for this observation

1.2.4 Packaging linguistic strings (and comprehensibility)

As discussed in Simon (1962), it is an inherent property of complex systems that larger wholes are 'packaged' in smaller units. As discussed in §4.1 constituent structure (i.e. organization according to clauses in language) is not a proof of biological hardwiring but an inevitable by-product of the problem at hand.

In principle, the either of the sentences in (4) could mean what the conventional sentence of (5) means.

(4) a. Jonathan and Catie sent a letter and a coconut to Lynn and David

b. Jonathan i Catie sent a letter i a coconut to Lynn i David

(5) Jonathan sent a letter to Lynn and Catie sent a coconut to David

The apparent efficiency of not having to repeat the verb may be outweighed by the need to untangle these two interleaved clauses. Nonetheless, *logically* this seems a possible configuration of syntax that (to my knowledge) it is not attested in human language. This in itself does not imply that there must be a stipulative system determining what can and cannot be part of language. Rather it is entirely possible that the gains of only saying the verb once are simply not worth the other complications. While it might be possible to understand if there were more indices in place, this too might well be a cost that is not worth the added effort when there is the simpler option of repeating the verb for each event described.

The question of what role the brain has in language can be addressed better by dividing the problem into a number of components. Without isolating these questions, innatists do themselves a disfavour in their attempt to attribute *every* aspect of human language to the uniqueness of the human brain. Primary among these questions is: what aspects of human language are inherent to *any* system of communication constrained by our articulatory and perceptual systems? Of more trivial interest (from scientific point of view) is which aspects may be the result of diachronic accident? While the effects of diachronic syntax are hard to judge and of less interest for the linguist-as-cognitive-scientist, they are unwise to ignore in principle. If we insist that our scientific model must incorporate these certain arbitrary facts, such as why adverbs appear before the verb in one language and after it in another, we risk distorting the basis of linguistic science.On the assumption that adverbs must be local to what they modify but that their actual position (before or after) is arbitrary, to try to provide a scientific explanation to this pattern (that may be due to diachronic accident) is a mistake.

The confirmation of any manifestly biological hypothesis must wait until we have a better understanding how the neocortex works (even those of other mammals or birds).

Therefore it is critical to narrow the realm of inquiry to that which is open to other explanations. Too readily do theorists attempt to assign every aspect of human language a reflex in the brain[11].

1.3. Modelling Birds or Building Airplanes

Many contemporary linguists and psychologists object to any attempt to derive linguistic ability from something like general cognition. Unfortunately the debate frequently become polemic and the rhetoric obscures certain fine points. Within the scientific domain, there is little doubt that the human mind has some degree of sophistication that allows it to use human language and clearly this sophistication is rooted in biology and seems unique to humans. These points can hardly be contested. The corollary, that there is an area of the brain that is *specialized for language*, does not follow. Unfortunately, claims for the biological hypothesis often claim that opponents of the Universal Grammar (UG) hypothesis are under some confusion about genetic determination. Popular discussion by Chomskyans (e.g. Boeckx and Hornstein (2003)) maintain this hypothesis but it is also found in works in related domains such as Pinker (1994) and epitomized by the following passage (Marcus, 2004, 86)[12]:

> It is popular in some quarters to claim that the human brain is largely unstructured at birth; it is tempting to believe that our minds float free of our genomes. But such beliefs are completely at odds with everything that scientists have learned in molecular biology over the last decade. Rather than leaving everything to chance or the vicissitudes of experience, nature has taken everything it has developed for growing the body and put it towards the problem of growing the brain. From cell division to cell differentiation, every process that is used in the development of the body is also used in the development of the brain. Genes do for the brain the same things as they do for the rest of the body: they guide the fates of cells by guiding the production of proteins within those cells. The one thing that is truly special about the development of the brain—the physical basis of the mind—is its "wiring", the critical connections between neurons, but even there, as we will see in the next chapter, genes play a critical role.

[11] A telling example of this is the Speech Act Phrase proposed to account for the condition seemingly present in first and second person pronouns because there are no inanimate pronouns, (Rizzi, 1997); Rivero (1994), Cinque (1999); as described in Tenny and Speas (2004): "there is a speech act argument (speaker/hearer) associated with the SAP"

[12] I do not quote this author gratuitously: Noam Chomsky is quoted on the book jacket as saying "*The Birth of the Mind* is a wonderful contribution to our understanding of the biological basis for higher mental processes. It unravels dilemmas, perplexities, and confusions, and carries the reader to the edge of current knowledge in areas of great fascination and promise"

> This idea that the brain might be assembled in much the same way as the rest of the body—on the basis of the action of thousands of autonomous but interacting genes (shaped by natural selection)—is an anathema to our deeply held feeling that our minds are special, somehow separate from the material world. Yet at the same time, it is a continuation, perhaps the culmination, of a long trend, a growing-up for the human species that for too long has overestimated its own centrality in the universe. Copernicus showed us that our planet is not the center of the universe. William Harvey showed that our heart is a mechanical pump. John Dalton and the 19th century chemists showed that our bodies are, like all other matter, made up of atoms. Watson and Crick showed us how genes emerged from chains of carbon, hydrogen, oxygen, nitrogen and phosphorus. In the 1990s, the Decade of the Brain, cognitive neuroscientists showed that our minds are the product of our brains. Early returns from this century are showing that the mechanisms that build our brains are just a special case of the mechanisms that build the rest of our body. The initial structure of the mind, like the initial structure of the rest of the body, is a product of our genes.

What is significant about this quotation is that it assumes that opponents of a 'mental organ' hypothesis[13] are necessarily opponents of any exact physical representations of brains and that to not support the innatist hypothesis in some shape or form is to be scientifically uninformed or a subscriber to 'deeply held feelings that our minds are special'[14].

The thesis of Marcus (2004) is essentially that that the brain cannot be seen as something separate from the body and is as much 'built' by the genes, as any component of the body. There can be little serious objection to this thesis. A false dichotomy has been created, assuming that everyone who opposes the 'mental organ' hypothesis necessarily argues against a detailed cellular structure of the brain. On the Mountcastle hypothesis, the brain has a very certain structure; the hypothesis is simply that the various areas of the neocortex (visual, auditory etc.) are less different from each other—possibly only trivially so—than generally presumed. That is to say, it is doubtful that whatever makes up the linguistic regions of the neocortex is specialized to the degree tacitly or implicitly supposed by most

[13]The exact terminology here is likely to raise arguments about details that nobody now knows; the basic division under discussion is that of the highly differentiated brain hypothesis, under which I include a 'language organ' hypothesis and those who believe it is more plausible that the neocortex has a common function to process *information* and that there may be little or no important structural differences between regions attributed to sight or to language processing.

[14]This is what is referred to in rhetoric as a 'straw-man' argument, attributing an easily destroyed belief or argument to an opponent in order to show the absurdity of this line of thought. This is unfortunate rhetoric but not unique in the literature; for example, in Anderson and Lightfoot (1999) they state that "It is clear that the body is not made up of cream cheese, and the same seems to be true of the brain."

articles in journals such as *Linguistic Inquiry* unless such a structure can be found in regions such as the visual cortex. A further question must be considered in passing regardless of whether the regions are differentiated from each other or not. If the brain needs something like a verbal node to recognize and process verbs, perhaps it also needs a *facial node* to recognize and process faces. The level of specialization this implies is problematic on many levels. Nonetheless this is not a trivial point but raises questions of how the brain might recognize *anything*; while it is conceivable that we have innate knowledge of faces, verbs, horses and trees, one has to wonder at what point this process would fail (how do we recognize cars, bicycles, telephones and umbrellas?) and what evolutionary process might underlie such rich specialization.

As far as talking about language as a result of biological specialization, Chomsky himself states that: "That is the property of discrete infinity. This property is virtually unknown in the biological world. There are plenty of continuous systems, plenty of finite systems but try to find a system of discrete infinity!" (Chomsky (2000a))[15] . Would this in and of itself not be cause to doubt the claims made for the biological basis of language? Language may be a *system nested in the biology of the human brain* but that is distinct from the assumption that it is a direct manifestation of biological specialization of the human brain.

1.3.1 The question of evolution

A further question, perhaps forever unanswerable, is that of a tenable theory of how the human capacity for language might have come about. While this is normally relegated to the sidelines—as perhaps is inevitable for a question with little likelihood of being answered—it should be addressed in any considered discussion of first principles.

It is worth noting that, in the years following Darwin's 1859 publication of *On the Origin of Species*, one of the most prominent critics of Darwin's theories was the Oxford linguist Max Müller, famous for coining mocking nicknames such as 'bow wow' and 'ding dong' for contemporary proto-language theories that attempted to derive language with reference to Darwinism (Fitch (2005)). Müller's objection to Darwin's theories was that because there was no known precursor to language in related species, Darwin's theory of evolution—while possible for animals—was not adequate to explain human evolution, particularly how humans acquired language.

Current discussions of the human capacity for language generally assume that, while language must be a product of our genetic inheritance, its lack of precursors is not a critical problem. Some theorists, such as Piatelli-Palmarini and Uriagereka (2005), try to find adequate motivation for a genetic 'great leap forward' (in their case, a possible virus that rewrote our DNA). Other suggestions, such as in Hauser et al. (2002a) (sometimes consid-

[15]Though of course one might look at this claim considering non-biological phenomena such as fractals. Biologists also state that some systems of single cell organisms act in some ways like one larger organism.

ered as Chomsky's aberration or recantation), try to minimize the biological distinctions between other animals and humans to recursion, a difference between what they call the faculty for language in the narrow sense (FLN), contrasted with the faculty for language in the broad sense (FLB). Obvious problems with this include the fact that a number of animals have been shown to have the ability for recursion (Marcus, p.c. REF) as well as the fact that, on the face of it, it throws out many of the observations about linguistic data that have been accumulated over the course of the generative programme *without accounting for the patterning of the data.*

The question necessarily arises whether this is the wrong sort of question and whether language is something more like an emergent property of the brain. If we can surrender our need to be distinct in *kind* and consider that our ability for greater recursion and greater pattern recognition and permutation may in fact be only a difference of *degree*, we would have a more plausible story from an evolutionary point of view. It also is important to note that it is often the case in nature that differences of degree lead to apparent differences of kind. Clearly, to speak or not speak is itself a difference of kind. Nonetheless, it does not follow that this difference of kind is proof of an underlying difference of kind or whether our relatively advanced abilities in recursion, coupled with our articulatory and perceptual systems lead to our basic abilities for language. These factors are again limited by the *impoverishment* of spoken language (no indices, no category identifiers), which are perhaps a direct result of limitation in our articulatory or perceptual abilities. The present thesis is that, once all these factors have been weighed and analysed, we will have a tenable theory of language that is consistent with general evolutionary principles, plausibly uniting Darwin with linguistic theory without need of any *great leap forward.*

1.3.2 Mapping vs. Specialization

Even if we make it clear that a theory of general cognition is not a fuzzy-thinking attempt at recovering notions of 'mind' but can be an articulated and scientific theory of the neocortex, many psycholinguists and psychologists might object to Mountcastle's hypothesis. Through such techniques as fMRI scans there is a great deal presently known about the different functions carried out by different areas of the brain and that damage to parts of the neocortex such as Broca's area or Wernicke's region have been demonstrated to affect language use. The inquiry into which regions of the neocortex are activated by what stimuli seems to reflect in many ways the 'language organ' hypothesis proposed in Chomsky (1965), that it is possible that there are 'organs' in the brain analogous to organs in the body. In certain ways these investigations have been fruitful; in other ways, if we take the evidence provided by Mountcastle, they have perhaps been misguided.

To resolve the apparent conflict in these findings, it helps to make the distinction be-

tween cortical specialization and mapping to an area of the neocortex. That is to say, are the areas in question related to various processing tasks because they are specialized for them or because the relevant sensory or motor neurons are *mapped* to them? To make a crude analogy[16], is the brain like a building with specialized laboratories in different rooms or is the brain like a library, specialized only for information storage and retrieval but without significant specialization from room to room? Continuing the analogy, if a stroke is like a fire in this building, will the destruction of room 4b destroy the fine arts department and the destruction of room 3a the biology books? Or would the destruction of these rooms mean that the photo dark room or the biology lab has been destroyed? That is to say, can the evidence for regions of the neocortex affecting certain functions be accounted for by the wiring or what it is wired to?

Evidence from the plasticity of the neocortex suggests that it is possible to consider the possibility that the neocortex has a common function from region to region. This has a number of advantages for neurological theory, possibly allowing us to unite findings in visual recognition with abilities for analogy in areas of language. Beyond this, it opens the possibility of a tenable theory for the evolution of the human brain from our non-linguistic ancestors, something that seems unfathomable under a language organ hypothesis[17].

1.3.3 Problems with Modelling

The most important question for linguists is whether language data can be accounted for as a function of general cognition. Nonetheless, it is worth discussing first some problems with modelling data without accounting for how the theory and the physiological evidence work together.

1.3.4 Cleaning up the data (and degraded judgements)

In order to begin addressing linguistic data from a point of view of logical analysis, Chomsky proposed that native speakers of a language have an innate knowledge of whether a given sentence is a licit sentence in their language or not. He proposed that this innate sense *must* be the result of the processing function of the language area of the brain. This supposition is a *hypothesis*[18] and one in need of less circumstantial evidence than has been

[16] Advocates of the 'language organ' hypothesis might argue that the idea is more metaphorical than literal, that it is a discrete entity in the brain without having a particular or single location; while this is true, the point I am making is unaffected: is there significant specialization from area to area or not?

[17] See, for example, the rather fanciful suggestions proposed in "The Immune Syntax: The Evolution of the Language Virus" (Piatelli-Palmarini and Uriagereka (2005)) that the necessary great leap forward allowing language to develop in the human brain might have been caused by a viral infection that altered human DNA; again I do not cite this randomly: the paper acknowledges feedback (and perhaps at least tacit support) by people such as Thomas G. Bever, Noam Chomsky, Giorgio Graffi, Margaret Kidwell, Andrea Moro and Donata Vercelli

[18] Except that for some researchers it "is not really a hypothesis. Rather, it is an empirical conclusion" (Legate and Yang, 2002, 151); this view seems more widespread than one would wish to see in what purports to call itself an empirical science; it seems that the principle of "working hypothesis" is itself poorly under-

found to date. In fact, it is entirely possible that user judgements are made from combination of 1) possibility of parsing the sentence and 2) analogy or reference to previously heard tokens. That is to say, the principle of speaker judgements was introduced with little empirical evidence and, as far as I know, has received none since.

What this principle has allowed in fact is for theorists to propose problems and possible solutions based on the judgements they have assigned to various sentences[19]. While in some cases this methodology is sound, particularly when the issue is concerned with interpretability of indices etc., in many other cases tokens are marked because the tokens are not deemed to be typical specimens of the language. While it is clear that language does not exist in isolation, the practice of soliciting judgements for tokens out of context is not generally considered problematic. In fact the discussion in Chomsky (1965) §4 would suggest that, without context, the full implication of phrase structure cannot be known. The example presented is as in (6).

(6) I had a book stolen

This is presented as a (possible) fragment that, with context, can mean anything from 'a book was stolen from me' to 'I paid for a book to be stolen' to 'I (almost) had a book stolen but they caught me' (21–22).

If the practice of marking certain features as ±[feature] has any typological or scientific merit, one might use this to better diagnose grammaticality judgements, perhaps in the following way:

(7) ±familiar; ±parseable[20]

a. +familiar; +parseable = always grammatical

b. −familiar; +parseable = grammatical with the right context or upon multiple readings or hearings; or that which seems too 'poetic' for standard dialects but that is not ungrammatical *per se*

c. +familiar; −parseable = fixed phrases and idioms

d. −familiar; −parseable = always ungrammatical or uninterpretable

The problem with such a system is that it admits that *familiarity* or context is enough to change the grammatical judgement. Therefore I question the validity of the notion of grammaticality as it is commonly used and suggest it is better described as 'interpretability' and 'familiarity'. As unusual as a phrase might be, if it is possible to parse it with a rich enough

stood; Chomsky himself has said that like any scientific hypothesis, it is refutable: "An innatist hypothesis is a refutable hypothesis" (Chomsky and Piaget, 1980 (1975, 80)

[19]The practice of arguing that, while one token may seem bad it is at least relatively better than another is particularly problematic, examples include Nissenbaum (2000) and Fox (2000)

[20]i.e. capable of being parsed

context and it is possible to understand it what is left is whether or not that configuration is familiar to the listener. That familiarity should be the basis of grammaticality judgements is decidedly less desirable. Nonetheless many tokens are judged 'ungrammatical' because the string of lexical items is unfamiliar in a particular series; if a sentence is comparatively *stranger* it may take a moment to place it in an appropriate context. It is certainly enough to make a phrase clear or interpretable to a listener but how can we understand this in the context of a discrete syntax processing device? Should something designed to derive and interpret syntax according to various agreement and co-reference relations be affected by *context*?

As suggestive evidence that there is no such processing device, I point to the phenomenon where professional linguists apparently lose their ability to make fine grammaticality judgements. That the ability to make these judgements can be degraded under a system of discrete computation is decidedly unexpected. On the contrary, this gives more credence to my proposal that sentences are allowed if there is an analogous or exact instance of that sentence in the linguistic memory of the speaker; professional linguists are perhaps somewhat unique as speakers of a language who are asked to seriously consider abnormal or even illicit tokens on a regular basis. On a mnemonically based system, it is unsurprising that this should degrade their judgements but that a 'syntax machine' would be affected by context and the unfamiliarity (or strangeness) of a token seems considerably less likely. I would expect a dedicated syntax processing unit to be oblivious to such extraneous factors.

Nonetheless, this working method is very prevalent among GB and MP linguists to the point that it is hard to conceive of working without it. It has been thought that in order to understand how the brain is processing language we must build a model of the grammar to 'see' what is happening. This leads us to a problem with another standard working method, the idea of discrete landing sites for movement.

1.3.5 Problems with Notation and Model Building

Following proposals that human language owes its commonalities more to restrictions in a hypothetical 'language organ' than to the logic of an adaptive communication system, Chomskyan linguistics is very concerned with mapping out how the data is manifested (abstractly or otherwise) within such a language organ. This practice in itself carries many suppositions and associated problems. It is tempting to accept the first arguments for transformational grammar on an intuitive level. Language *does* seem compositional. It seems reasonable that rather than memorize all the forms of syntax that a speaker might encounter, we have some way of moving words around to change a statement to a question. In sentences as in (8) and (9), it seems probable that the second sentence is derived from the first.

(8) a. This boy can solve the problem

b. Can this boy solve the problem?

Likewise, if words have such regular patterns—in English at least transitive verbs need an object and that object normally follows the verb—then we might presume that the verb in (9) at some level needs an object to follow it.

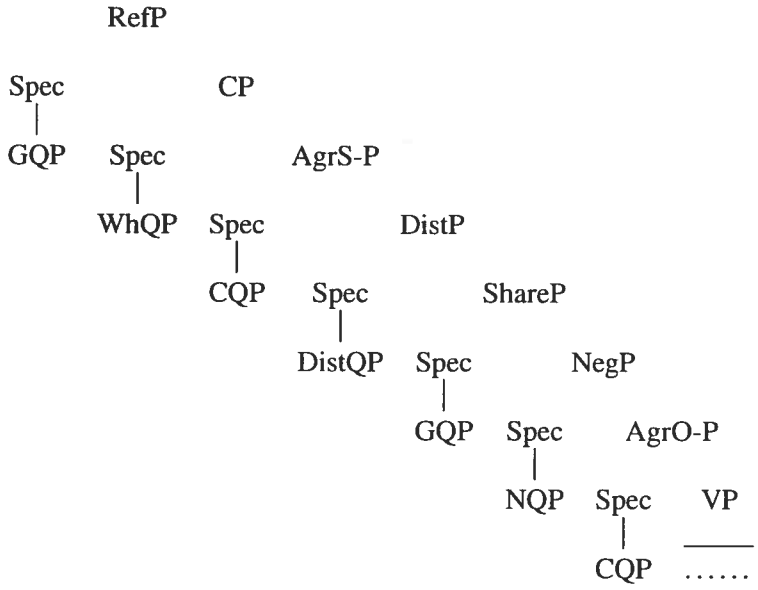
(9) a. The girl loves x

b. The girl loves who?

c. Who does the girl love?

At one level the argument is very compelling; it is hard to deny that there is some level of compositionality involved in human language. The question is, can this be modelled in any insightful way? To assume that predictable syntactic configurations imply or necessitate some kind of derivational or representational structure (as in the tree structure in (10) leads to a number of conceptual problems[21].

(10) From Beghelli and Stowell (1997)



Just as a chemist does not predict that all molecules are underlyingly the same molecule and that a carbon atom (with six protons) should somehow carry the 'landing sites' for larger

[21]This tree structure can be considered an extreme result of following comparative modelling to its logical conclusion. The explanation that Beghelli and Stowell offer for these facts is based on multiple landing sites for the quantifiers, corresponding to the featural specifications of the QNPs. The proposal in Beghelli and Stowell (1997) is basically as follows. While 'each' and 'every' are both distributive QNPs, only 'each' is obligatorily so. Both 'every' and 'each' can access spec,DistP because they are singular (contra 'all') and as such are available to the $Dist^0$ operator. The two QNPs vary in terms of their specification for the [±Distributive] feature: 'each' is [+Distributive] while 'every' is underspecified for the feature. As such 'every' can only move to spec,DistP when their "set variable is not bound by a lower operator such as negation" (Beghelli and Stowell, 1997, 103-104).

atoms even if some reaction cause the one to become the other. Works such as Chomsky (1995a) have previously argued against what has been perceived as theory internal artefacts but have contented themselves with details of the notation. The notation is important because it becomes the model with which we envision what we cannot see; poor notation can lead to poor understanding of the problem. We cannot look into a sentence or into the mind of the speaker and see whether or not there are unused nodes for possible adjunction or grammatical forms or even know if our model is on the right track. Since Pollock (1989) comparing the way one language works with another has been considered one of the best ways to make serious theoretical generalizations. This has seemed to be a tremendously important tool in the study of language in order to regularize predictions. Nonetheless, regularizing predictions to the point of (analogously) creating hypothetical carbon atom containing 92 atomic landing sites because a uranium atom has this many protons in it does no favour to our investigation.

Our ideas about structure are based, as Cinque's hierarchy, on comparative modelling. Yet, just as chemists do not suggest that the uranium molecule and the helium molecule have the same number of possible adjunction sites, so too linguists should not suggest that a given sentence or a given language has the same number of possible adjunction sites because another one has it. The chemist knows that there are properties of the materials that can be generalized as a *system* and that these allow for chemical 'composition' if all factors are correct. So too, we should know that the *system* of language allows for further composition if all factors are correct. Just as chemists had to disabuse the public about the Mechano model of chemistry (the sticks between the atoms), it is important that linguists do not take comparative models (such as X-bar trees) so seriously that they arrive at seemingly rigourous but empirically dubious models (such as Kayne (1994)).

Inherent in any difficult modelling problem is the possibility of creating a logically consistent representation of the data that is completely at odds with the subject that is being modelled. This problem is stressed in the following quotation from Michael C. Reed, a mathematician who studies biological systems:

> The second difficulty is that a priori reasoning is frequently misleading. By "a priori reasoning" I mean thinking how we would design a mechanism to accomplish a particular task. As a simple example, both birds and planes burn something to create energy that can be turned into potential energy and both have to use the properties of fluids that are implicit in the Navier-Stokes equations. But that doesn't mean that one understands birds if one understands planes. To understand how a bird flies, one has to study the bird. Modelers are sometimes satisfied that they have created a mathematical model that

> "captures" the biological behavior. But that is not enough. Our purpose is to understand how the biological mechanisms give rise to the biological behavior. Since these biological mechanisms have been "designed" by evolution, they are often complicated, subtle, and very special or unusual. To understand them one must immerse oneself in the messy, complex details of the biology, that is, you must work directly with biologists. ("Why is Mathematical Biology so Hard?" Michael C. Reed Department of Mathematics Duke University)

That is to say, theorists are inclined to believe that the theories should compete with other theories and refine themselves and in the end we can be assured that the theory that best accounts for the data will be right. Because of the nature of the problem, even if there were a theory that accounted for every last bit of data without exception, we would still not have any assurance that we would have a model of what goes on in the brain. In the mainstream of linguistic theory, it is the theorists who dominate the stage; too often experimentalists and acquisitionists are working in the service of theorists who are very disconnected from the nature of the problem at hand. The problem is how can a speaker communicate to a listener in a system that quite likely prioritizes efficiency over displays of complexity or 'beauty'? If we find Chomsky's recent musings that language must have 'evolved' (biologically) for beauty somehow unlikely, then we might consider that it evolved *as a system* for communication. What then, is the nature of the problem and the most fundamental limitations and character of this?

Chapter 2

Postulates

In this chapter I briefly review the base assumptions of my analysis. Although some of these may seem variably controversial, trivial or overly evident, because I am re-evaluating a number of base assumptions of the Chomskyan system, I cannot proceed without determining in some detail what seems to be necessary in a scientific programme investigating human language.

2.1. The nature of the problem

The task of the linguist studying syntax is to understand what factors affect the syntactic structure of the language. To ignore how syntax is used and the conditions of the transfer of information between speaker and listener is to overlook one of its most basic conditions[1]. Most human language is communicated as sounds directly conditioned by the human articulatory and perceptual systems. To communicate anything but the simplest thought, we must use a series of sounds and these must be produced and organized over time so that these lexical items come one after the other. The practicality of such a system of sound organized in time necessitates that a speaker or listener can remember and comprehend this information stream, which may require some 'unpacking' or reorganization. Due perhaps to a variety of constraints (articulatory, perceptual) and perhaps expediency, language is not as richly detailed as might be possible (see §2.14). This lack of richness I refer to as the impoverishment of the information stream. I assume this to have a marked effect on the possible forms of syntax.

Furthermore, the building blocks of the structure in question (cf. chapter 3 and Baker (2003)), i.e. lexical categories and lexical items, derive from the need to establish *reference* (the basic role of nouns and pronouns) and indicate various details of the state of the

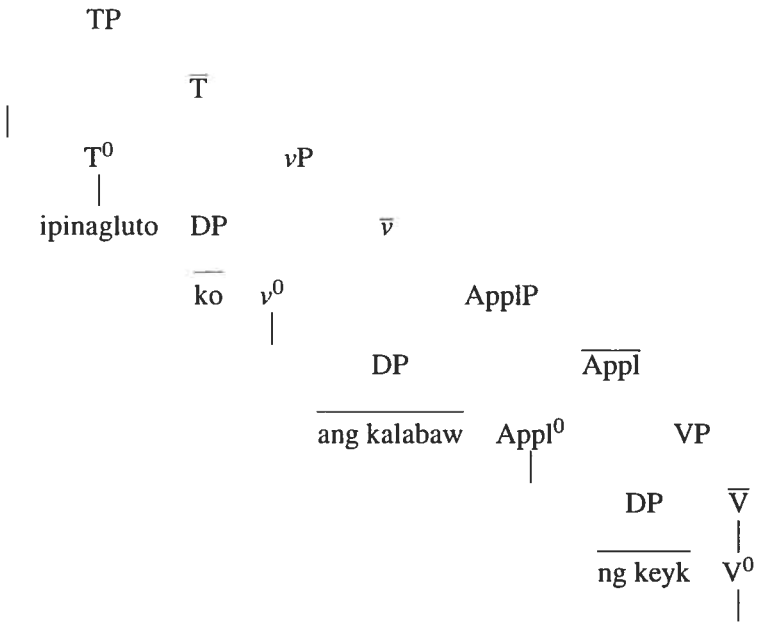
[1] Although some theoreticians might argue that syntax did not develop to transmit information, this seems to be how it is commonly used today. That is to say, in recent MIT lectures Chomsky has speculated that the faculty for language may have evolved for some kind of display of beauty, perhaps analogous to a peacock's feathers; nonetheless, to my knowledge, these speculations are not more than that. (The observation is nonetheless interesting in that it shows the problem of accounting for how such a system might have evolved.)

referents (dynamic, static, continuing, past etc.; the role of verbs). The nature of verbs is to condition the reference of nouns (see §2.7 and (12a)). This seems to require a consistent positional relation with the noun (or *wh*-word). Alternately the verb seems to be able to hold this positional relation with the elided *canonical position* the noun occupies in the simplest phrases. If nouns or *wh*-words are displaced (for various reasons including focus and question formation), they seem to need to have the capacity to be reconstructed into their canonical position. If the noun is not able to be unambiguously reconstructed to this canonical position (because another item of like category intervenes) intelligibility suffers. As such locality (i.e. uninterrupted proximity to relevant lexical items in a phrase) is an important factor in language.

2.2. Considering Occam's Razor and the principle of Economy

As in any scientific investigation, we must take Occam's Razor seriously and no more assumptions should be made than are needed. That is, our null hypothesis must follow principles of parsimony, that they behave in the simplest or most economical way. Nonetheless, what is the simplest or most economical is subject to various opinions and much debate. As discussed in Marcus (2004), organs such as the eye can take many forms and tend to have seemingly *inefficient* elements, such as the blind spot at the back of the mammalian eye where the retina is not uniform so that the optic nerve can attach to the eye. If we are given the task of designing a faculty for language (in a computer or synthetic animal for example), perhaps we might find it efficient to build discrete processing centres or 'nodes' as in (10) in §1.3 or (11).

(11) Gloss: 'I baked the water-buffalo a cake'[2]



Once the structure is in place and the corresponding items discretely identifiable, such a strictly serial system might well be the most 'efficient' system. One the other hand, if we are looking at efficiency in terms of evolutionary pathways (cf. §1.3.1), it is important to carefully consider our commonalities with our next closest relations and examine how the data (in this case linguistic) might be accounted for. Having a discrete processing unit that can easily and uniquely recognize items of a lexical category, e.g. a verb, and to know that it is intrinsic to this category that it needs to assign thematic roles and be 'saturated' by nouns is perhaps an extremely efficient or economical system. Nonetheless, just as mammalian eyes are sub-optimal (needing two to make up for the blind spot in each of them), it is reasonably clear that we cannot account for linguistic data with just any measure of efficiency. Although there are arguments in favour of different perspectives, following Mountcastle's hypothesis (and general evolutionary concerns), I favour efficiency in terms of evolutionary pathways over other possible efficiencies. This essay argues that the data does not preclude such considerations. (See also discussion in §4.6.)

2.2.1 Economy of scale and the burden of proof.

A central postulate of this essay is, as Chomsky himself stated, "the burden of proof is always on the proposal that the theory has to be made more complex"(Chomsky, 1995b, 241). As such, analogy can be made to the principle of *economy of scale*. By this I mean that it is preferable that the biological underpinnings of language can be limited to something

[2]Thanks to Norvin Richards for the structural representation and original sentence and Raph Mercado for confirming the Tagalog data. Note that order is irrelevant in this example; because Tagalog is a scrambling language, this data cannot be construed with any assurance of canonical position or 'deep structure' etc.

like general memory and general processing limitations (as well as limits on the information stream). If the brain has 'invested' in memory or a general ability of connection-making (or recursion), it is more likely that it will use these if it can. This is a choice between more than our aesthetics or principles of better design. Assume one theory that needs a creature to have evolved to have a lot of general memory and something to recursively combine elements from this memory. Compare this with another theory that needs a creature to have evolved to have this same type of memory (but perhaps less of it) but also needs a discrete computational system and a discrete lexicon and lexical categories and items with signatures identifying these lexical items to the computational component. I argue it is reasonable to assume the theory requiring more memory (possibly much much more) is better.

An example of this is that, while passive constructions have typically been considered likely candidates for transformations (Chomsky (1965)), it is quite likely that, in terms of cost to the system, it is 'cheaper' to memorize both active and passive patterns than to invest in a transformational mechanism of some sort.

2.3. Mountcastle's hypothesis should condition our theory

Language is arguably the richest source of output data produced by the human brain. As such, Chomskyan linguistics has traditionally been very interested in how the brain might condition language and in turn what language data might say about the brain. For this reason, linguists cannot ignore findings such as proposed in Mountcastle (1982) regarding the structure of the neocortex. Specific to Mountcastle's hypothesis, as discussed in §1.2.1, the burden of proof is on the theory advocating more specialization between each area of the neocortex.

This is not to simplify the nature of the neocortex but to argue that whatever it consists of, the function common to different areas (e.g. visual, auditory etc.) is more significant than that which distinguishes them, specifically the areas responsible for language. It is therefore imperative to condition our theorizing and interpretation of linguistic data so that the theory requires less specialization in the biology. While language is clearly a function of human biology, it is not clear whether it is a primary function with specific and discrete components specialized for the computation of language, or a secondary effect, a system functioning with the workings of the biology that may itself not be specialized for language processing per se. Is it worth asking if there is anything particularly *visual* about the visual cortex or *linguistic* about regions such as Wernicke's and Broca's areas? Very possibly the respective areas process *information* from the various sensory (etc.) inputs without necessarily needing to be specialized for the *type* of information. That there are differences in the nature of linguistic and visual information, I propose, derives from the nature of the information itself and does not necessitate a different neural function. Spoken syntax

encodes a stream of information and this stream is constrained by time. Such a system inherently has limitations and consequent forms caused by the nature of the medium (see §2.14). While the brain uses a lot of parallel processing—we can walk and talk without having to alternate between one and the other—we can only speak one word at a time. In order to encode information, such as reference, background and emphasis, syntax takes on many different forms that might not exist if it were in a medium with other constraints (CF. DETAILS OF ASL).

2.4. LF, QR, Distributivity and the interpretation of test results

The interpretation of test results is a critical matter: deciding what tests are valid, if the data is relevant and what the data implies. In generative grammar much is hypothesized to move covertly, without audible indication, presumably to rearrange so that a 'semantic component' of UG can process the syntactic string. The form it is thought to attain in order to be processed is called Logical Form (LF), and this level is generally considered to bear certain similarities to the formal logic developed by western philosophers and logicians (Hornstein (1995) and references within).

The question is when should we take the test results to suggest a significant reworking of grammar (into LF),and when should examine it and consider the test results to have other causes. Because current theory assumes an essentially serial system of syntactic processing, the parser cannot take advantage of linguistic objects further along in its trajectory; this is reinforced by certain theory-imposed conditions such as no 'look-ahead', which suggests a single, unidirectional pass of the parser. In addition to maintaining a strictly serial view of processing, this ban on the parser looking ahead of itself suggests a single pass of the parser; yet in in our machinery designed to read strings of data (taking the crude example of a laser reading a CD or DVD), a parser that can only pass once is prone to error and far from ideal. Given that we talk so slowly relative to how fast our brains apparently think, it seems reasonable that our linguistic parser could compute and re-compute a string, perhaps considering inverse orders of certain so-called 'logical operators', before reaching an interpretation.

As such, it is worth weighing different options that might explain the same data[3]. As discussed in better detail in §3.6, the primary theoretical justification for LF is quantifier raising (QR). QR is thought to be necessary to account for when a universal quantifier, $\forall$ (e.g. every), 'scopes' over an existential quantifier, $\exists$ (e.g. some). By 'scope' it is generally meant when the universal quantifier has a distributional interpretation over an existential quantifier; in a sentence such as "every boy loves some girl" the universal 'every' distributes over the existential 'some' if for every boy there is a distinct girl he loves. This contrasts

[3] At the very least, for the sake of evolutionary plausibility.

with the reading where 'every' does not distribute (or scope) over 'some' and every boy loves the same girl.

The question is, how do we account for this data? If the attested positions of the quantifiers are in the order $\exists > \forall$ but a distributed reading is available, do we build a model that presumes the quantifiers rearrange 'covertly' from their attested positions into a position such that the universal is linearly prior to the existential? Or do we suggest that the parser might be able to pass along the same string more than once, considering different configurations of the ordering or perhaps considering running itself in reverse for the purposes of these quantifiers? I suggest this is a simpler model that supposes less mental machinery. The other obvious advantage of this model is that prosody, sentential stress and other focus effects tend to disrupt quantifier interpretation (Herburger (2000)). A multiple parsing models seems more easily accountable to prosody, focus and pragmatic effects than one that presumes certain innate and discretely specialized quantifier 'raising' mechanisms.

A related matter pertains to the claim that there is such thing as a 'logical argument position'. Even professors of logic recognize that the ordering used in mathematical logic is determined more by convention than any inherent trait of the operators they have developed (James Loveys pc.). I suggest that we are better to talk about *canonical* argument positions and that these are largely conditioned by language-specific diachronic developments, in combination with principles of legibility suggested in this essay.

2.5. Universal principles of language are non-arbitrary

The basis of language is essentially non-arbitrary and in the sense that mathematics is non-arbitrary. That is, effects such as *wh*-islands derive from the restrictions caused by the medium (spoken and heard sound) and their basic form does not depend on our biology. In Chomskyan terms, these resulting non-arbitrary principles are roughly analogous to *principles* of Principles and Parameters theory.

Mathematics is the most obvious example of a non-arbitrary system. It is in principle possible to derive elementary algebra from the simple counting of whole numbers (as demonstrated in Feynman et al. (1963)). An illustration of a non-arbitrary system is elementary algebra. Once we allow the ability to count[4], this leads to addition; the opposite operation, subtraction, is quickly apparent. Subtraction then suggests the possibility of negative numbers (e.g. 1 minus 2). Multiplication can be described as repeated addition; division is the opposite operation. Exponents and logarithms follow from these same principles.

I argue that such phenomena as subjacency, barriers and phases can be derived from the

[4]Perhaps itself requiring a fair amount of evolution but an ability present in limited form by other animals; see work by Irene Pepperberg for instance.

limits of reference within a sentence due to such things as described in §2.14. If such things can be explained by the limit of reference (see also §2.11), then they can be considered as examples of non-arbitrary limits of human language.

2.6. Language-specific forms are a mix of arbitrary 'accident' and non-arbitrary principles

There are necessarily arbitrary elements to any specific language, perhaps due to diachronic developments. These developments include the form of lexical items (which are arbitrary in the sense of Saussure (1916)); syntactic arbitrariness includes such things as whether an adverb is normally before or after a verb. While in Principles & Parameters (P&P) theory (Chomsky (1981); Chomsky (1995b)) such arbitrariness is considered as something like a 'biological switch' (i.e. a *parameter*), it seems more plausible that these are typological generalizations and parameter-like consistency is due to the tendency of the speaker to be essentially conservative, repeating the same pattern unless a contrast is required (see §2.12).

2.7. Lexical categories contain no identifying 'signature'

I assume that any discrete computational function (of the brain or other computing system) must depend on some kind of identifying signature for the objects it manipulates. Suppose that a hypothetical discrete grammatical component were able to classify verbs according to the principle 'a verb is something that takes morphological agreement markers' e.g. on *I run, you run, she runs,* there logically would be some kind of phonologically null marker on the 1st and 2nd person singular forms of these verbs. Nonetheless, because these markers are phonologically null, they cannot be communicated to the listener. Therefore the grammar must be organized in some way so as not to need discrete reference to the lexical category of a lexical item.

Some lexical categories, such as regular verbs in many languages, can be 'tagged' by markers such as *-ed* in the English simple past tense. Nonetheless, these tags are highly irregular, inconsistently applied and generally not part of the root of the lexical item (REF?). Therefore, the tags that are manifestly attested do not seem sufficient for any hypothetical discrete computational system to identify them in a category in any reliable way. If there are tags that somehow covertly exist in the mind of the speaker, they do not seem to be part of the spoken syntax; as such these cannot be read by a listener.

An analogy to cell biology may be instructive here. Interactions between proteins are necessary for cells to carry out important functions. Interacting proteins are identifiable to one another based on their structures and ability to 'interlock'. Nonetheless, such metaphors are only useful if we keep issues of the 'cognition of the system' to a minimum. By this I mean that we do not want to invest any seemingly cognitive element in a system that is

itself part of cognition.

With this in mind, in chapter 3, I examine the definition of lexical categories. In Baker (2003), two definitions of verbs are offered. First they are described informally, according to how they relate to other lexical categories, as in (12a); Baker then proceeds to define them in terms of the structure they contain, as in (12b).

(12) a. verbs "are inherently unsaturated expressions that hold of something else, and thus the nucleus around which sentences are typically built"

b. X is a verb if and only if X is a lexical category and X has a specifier

(23).

Without a discrete system of grammar this second definition is not tenable. Nonetheless, *with* a discrete system of grammar the system only seems plausible so long as the representation of the verb remains in the mind of the speaker. Once it is spoken, there is no consistent signature available to the listener, i.e. depending on the person feature in question it may or may not be spoken. If it is not spoken, it is not communicated to the listeners hypothetical discrete system; therefore there is no reliable signature. It seems then that verbs at least must be identified, not by a category marker, but by position in the sentence and by speaker knowledge of the lexical item in question. If verbs are to move to check features (on the way to LF), it seems to be a fairly haphazard system to rely on speaker knowledge of a lexical item or position within the sentence.

This is not to deny the notion of lexical categories but there is no basis for assuming that they are something innately fixed in the brain (these like, e.g. multiplication, can be derived by our ability to refer (etc.) see §3). They exist in some functional sense, as in (12a), as well as in the analytic generalizations we can make when we study grammar. Just as it is highly unlikely that we have a special 'multiplication processor', the idea that there is something like 'features,' that a verb might be attracted to and move to check, is specious.

Furthermore, contra Chomsky (1972) and as discussed in detail in chapter 3, the lexical categories cannot be simplified into a binary system of ±Noun, ±Verb (Baker (2003)). Even if lexical categories did exist as essentially arbitrary items (perhaps supplied by our biological heritage), on the assumptions of §2.7, they cannot be expected to be identified by a speaker or learner without effort. Therefore even if something like a quantifier is provided by innate knowledge, the identification of it is a problem. Strange behaviours, such as described by the theory of quantifier raising, seem unlikely if the system has no clear means of identifying these categories in the first place.

2.8. Lexical items are acquired and not triggers for innate knowledge

Following from §2.7, lexical items themselves are much less precisely defined than many would like to believe. Chomsky states: "There is overwhelming reason to believe that concepts like, say, 'climb', 'chase', 'run', 'tree', 'book' and so on are fundamentally fixed [by our genetic inheritance]. They have extremely complex properties when you look at them. This is not recognized in traditional lexicography. When you read the huge Oxford English Dictionary (the one you read with a magnifying glass), you may think that you are getting the definition of a word but you're not. All you are getting is a few hints and then your innate knowledge is filling in all the details and you end up knowing what the word means" Chomsky (2000a) (but see §2.12.1 for more discussion). Nonetheless, even if this were so, it is of questionable merit to attempt to maintain our lay assumptions about the fixity of these concepts at the expense of feasible means of connecting them with the world[5].

This is not to deny that these lexical items may be very complex on analysis. It is simply to say that there is no basis for assuming that the definition we arrive at upon analysis is anything like the representation in our minds. We might simply refer to previous instances of trees or running when we learn the lexical item pertains to these.

Tokens of lexical items like 'climb', 'chase', 'run', 'tree' in a given language clearly have audible and detectable signatures[6]. These are what allow us to recognize and use them in sentences. The associations and representations created in the mind form the basis of what we call the 'lexicon'; the element that seems to be missing is how the audible signatures of these lexical items (for the most part developed diachronically, with any given phonological signifier of a 'semantic primative' developing along often quite distinct paths) can be identified discretely even though they lack a particular tag that might identify them as some archetypical, biologically endowed concept in the brain.

---

[5]Furthermore, if we have an innate understanding of, say, 'climb', 'chase', 'run', 'tree' then do other animals as well? Does a dog or a squirrel or a sparrow have some understanding of what they are doing when they are engaged in these activities or when they see or climb the tree? We might presume that animals lack the awareness of self and environment that humans have but it is clear that they can recognize other members of their species (a dog can recognize another dog) as well as those species that are their predators or prey; likewise they can recognize when they are running or running up a tree. This is not to presume that they need to have as nuanced a description as we might. Nonetheless, if we presume that we need innate representations of, for example, a dog in order to recognize a dog, this suggests that, if a dog can also reliably recognize another dog then it too would need an innate representation. This reduction can continue to the point of absurdity; any creature that needs to reproduce sexually (i.e. recognize another example of itself, and one of the opposite sex) or that preys on, or is preyed upon by specific species, would need a mechanism to identify these other creatures. It is somewhat absurd to assume that in any creature that is at all sentient (i.e. perhaps including insects but excluding less sophisticated creatures) needs a fixed representation of that which concerns it.

[6]This is not a different sort of signature than discussed in §2.7 but rather a subset of all possible signatures, i.e. in §2.7 and these correspond to the phonological identify of the given lexical item. I pointed out that the system is in need of a consistent means of identifying such things as syntactic category if categories are how the grammar is organized; a given word ('chien') in a given language (French) clearly has a signature so we can relate sound and meaning.

Lastly there remains the open question of why linguistic universals (prepositions, lexical items etc.) are not more regular if biologically determined. A theory of 'parameter switches', if otherwise tenable[7], might explain this but must answer why simpler explanations (e.g. exposure) does not answer this better. The discussion in Wittgenstein (1958) seems a more pragmatic approach (comparable with the suspicion expressed in §3)[8]. There Wittgenstein asks what we need to know about a word so that its use does not produce argument. Contra the long history discussed in Chomsky (1965) going back to at least Plato of answering that this knowledge must be full and innate, Wittgenstein suggests that this minimum knowledge, of perhaps the characteristics of a chair, dog or a tree, may be derived from the limited knowledge (and subsequent generalization) of a limited number of instances of these items. Saul Kripke's explanation in 'Wittgenstein on Rules and Private Language' helps explain often confused ideas about Wittgenstein's 'language game' when it comes to the lexicon:

> On Wittgenstein's conception... We cannot say that we all respond as we do to '68+57' because we all grasp the concept of addition in the same way, that we share common responses to particular addition problems because we share a common concept of addition.... For Wittgenstein, an 'explanation' of this kind ignores his treatment of the sceptical paradox and its solution. There is no objective fact—that we all mean addition by '+', or even that a given individual does—that explains our agreement in particular cases. Rather our license to say of each other that we mean addition by '+' is part of a 'language game' that sustains itself only because of the brute fact that we generally agree. (Nothing about 'grasping concepts' guarantees that it will not break down tomorrow.) (Kripke, 1982, 97)

That is, as pointed out in §2.9, the analysis of many instances of a word by an academic may well not correspond to the way we use words and the way the brain categorizes information pertaining to them[9].

---

[7]While is is fully conceivable that something like parameter switches might exist, this neither means it is necessary or the more likely explanation. Physiological evidence needs to be supplied as well as evolutionary plausibility.

[8]This is not to suggest that I arbitrarily accept other aspects of Wittgenstein's views on language; his views do not address questions of such things as *wh*-movement limitations and adjuncts—questions Chomsky is right to point out must have another cause.

[9]An interesting insight into this question is the nature of Semantic Dementia, where, as explained by University of Manchester researcher Lambon Ralph, "For these patients, it's not like words have been deleted from the dictionary, so that you know about a duck one day and then the next day you don't ... Instead, your information about ducks gradually gets fuzzy, and so you sort of vaguely know what a duck's like, but you don't know the details" (from BBC website "Strange ducks shape brain science" (Wednesday, 6 September 2006). Such cases of lexical breakdown do not seem characteristic of discrete systems.

2.9. Lexical items/categories and linguistic tokens are recognized with reference to prior exposure

If a hypothetical discrete computational component of grammar could not recognize and process lexical items and categories, the organizing principle of grammar needs some other mechanism to produce language. Such a system may very well be a robust system of what we can loosely call memory, i.e. an input/output system that, on a lexical level, recognizes previous occurrences of a phonetic pattern (the token) and generalizes a phonological signature from this so as to connect new tokens with old; on a syntactic level, the mental parser needs to be able to recognize and generalize patterns so as to recombine them in similar phrasal tokens. Under this assumption, the tokens can be related to other tokens (whether on the lexical word level or syntactic level) and presumably over time build up a reference. This reference would not be to archetypical types from our genetic inheritance but generalizations with particular audible signatures for the *signifiant* and perhaps visual and other signatures for the *signifié* (see also §2.12.1).

These signatures for the *signifié* might be very much the same thing as generalizations of them, i.e. based on (many) interactions (etc.) with the thing itself. This assumes that just as, with repeated exposure, generalizations of a sound develops into a phonological representation of a 'lexical item' so too do generalizations of visual information (etc.) develop and become associated with such a representation; this hypothesis assumes that two (or more) such elements work together to create a general mental representation of an item in the world without necessarily having reference to a tidy pre-existing archetype. The details that build this representation are the various 'tokens' (whether 'tokens of sound' or 'tokens of experience') that went into building it (cf. work by Joan Bybee, e.g. Bybee (1995)). Because of the very nature of how these representations are created, any academic analysis is necessarily very complex (see the lexicographical work by Igor Mel'cuk for example, e.g. Apresjan et al. (1969), Mel'cuk (1988)).

2.10. The serial nature of spoken language is a constraint on output, not assembly

From the beginnings of Chomskyan theory, there has been a tendency to assume a serial interpretation of syntax and the classic argument for the innatist hypothesis relies on this assumption. (See §4.5 for discussion.) More recent theories such as Richard Kaynes work in 'ambiguous paths' (Kayne (1984)) and his Linear Correspondence Axiom (LCA) from Kayne (1994) also suppose a strictly serial syntactic parser. Unfortunately, while this work is interesting from a purely abstract point of view, it is entirely dependent on an innate structural representation and prone to certain theory internal problems. As discussed in §2.2, an ideal system might well be a discrete, genetically-determined serial system because, if sta-

ble and meeting other requirements, this may be considered more efficient. Nonetheless, problems for a consistent structural description of sentences (e.g. adjuncts vs. complements in X-bar syntax), barriers to movement within the structure (e.g. clausal subject islands) and the interrelation between various components of the structure (e.g. antecedent contained deletion/ellipsis) suggest that a serial understanding of the structure is riddled with conceptual problems. Even if the structure itself is considered to be constructed by apparent 'non-linear' means, typical representations in syntactic structural 'trees' are completely serial. Our experience of sound_x being expressed in time_y perhaps leads us to this way of thinking.

As discussed in Gibson and Pearlmutter (2000), it seems correct to assume that the means by which we analyse language with a number of parallel—or at least *non-serial* operations, i.e. although we speak and hear words in a series, there is little reason to believe our comprehension of language is necessarily serial. While Gibson and Pearlmutter (2000) look at the question from an experimental point of view and look at reaction times between tokens (arguing for a ranked, parallel parser), I reanalyse certain key structures from a theoretical perspective; this reanalysis, particularly argument-adjunct interactions, is central to this essay.

2.10.1 Unacknowledged non-Serial theories

If there is empirical evidence for the non-serial composition of phrases, one might expect there to be implications in another theory studying the same empirical data. While generative grammar maintains a general assumption of serial structure, the empirical evidence for the non-serial production of sentences has been accumulating within the theory. Examples include Lebeaux (1988)'s Late Merger of Adjuncts (LMA), the Path Containment Condition (PCC) of Pesetsky (1982) (also see May (1985)) and Richards (2001)'s Principle of Minimal Compliance (PMC). The principles I maintain are implicit to non-serial phrasal production are that the production of adjuncts takes place separately and assembly does not take place until shortly prior to speech. That one phrase cannot 'see' into another is suggested by the Path Containment Condition; that these components are interleaved separately is suggested by the Late Merger of Adjuncts theory; that there is a relation between 'not seeing into' a contained clause and it being interleaved is suggested by the Principle of Minimal Compliance. See §4.11.

2.11. Locality derives from the impoverishment of the syntax

A central underpinning of this essay is that, following from §2.7, locality derives from the impoverishment of the syntax and functions as the weakest condition of association. If no overt reference is made between two lexical items in a syntactic string, locality between

them may suffice as an implied connection. This is a very weak condition of association but, in the absence of more overt connections between two items, perhaps perfectly clear. If the speaker can save effort by not pronouncing extra morphology (to function as indices) when locality achieves the same result, perhaps this is a form of efficiency.

Comparison can be made with other systems, such as cellular telephone network towers and client (i.e. 'user') handsets. In this example, we must ignore the extrinsic issue of allowing human communication and focus on the intrinsic issue of the cellular tower communicating with the handset. What is important is that the cellular tower must constantly know that it is in communication with the client. This technology depends on being able to recognize the clients in a discrete fashion. For cellular telephone communication, the tower must constantly be communicating with the handset, checking that the user is the same user as it was fifteen seconds previously, perhaps displaced by a metre or one hundred metres from the previous location; if the user with the handset moves too far, to the edge and beyond of one tower's range, the system must have a means of 'handing off' the user's call to the next tower in the network so that the new tower can distinctly and uniquely identify the user and not 'drop' the call. If the tower or server cannot discretely recognize the user's interface (client handset with a distinct identifier), they cannot transfer data with the knowledge from one moment to the next that it is the same user receiving it and therefore making sense of it.

While the metaphor is not perfect, analogy can be made to the system of language. If all components of a system (in this case, a sentence) are recognizably distinct, then it may be no problem for the system to 'hand off' one WH word to a position further down the line. If, following Rizzi (1990), interveners of a like (and recognizably so) category come between one otherwise recognizable item and another, this interference may cause the system to 'drop' the signal (i.e., cause an uninterpretable sentence). The system may therefore limit itself to situations where there are no intervening like categories, as described by Rizzi's *Relativised Minimality*. A direct consequence is the common side effect of locality; this is comparable in this metaphor to the characteristics of a citizen's band (CB) radio, where there are no discrete systemic identifiers and the system is forced into a very definite (but not discretely so) system of locality because otherwise interference would cause the system to be unintelligible.

This has been stated as *Shortest* in Richards (2001):

(13) Shortest

A dependency between the members of a pair P of objects $\{\alpha, \beta\}$, where α and β are (possibly singleton) sets of coindexed[10] elements, obeys Shortest iff no well-

[10]Richard's footnote: "By 'coindexed' I mean simply 'belonging to the same chain'; no claims about the

> formed dependency could be created between the members of a pair P′, created by substituting γ for either α or β, such that the set of nodes c-commanded by one element of P′ and dominating the other is smaller than the set of nodes c-commanded by one element of P and dominating the other.

The point to be made here is that, while this definition of locality is structurally defined (by mention of c-command, chains, nodes, etc.), the principle behind it is not dependent on either structure or the innate biology supposed for the structure. In this sense, locality is what comes as the default organizational system if a more sophisticated system of identifiers either breaks down or is otherwise not used. This is to say, while locality may be considered as a fairly arbitrary stipulation of the grammar—or even of any system of organization, it is important to analyse it for basic causes. My insistence that locality has non-innate roots is, in fact, not out of step with certain current trends in the field. In his 2004 LSA address, Chomsky compares earlier generative work with recent Minimalist Program (MP) ambitions. He recounts that UG principles were previously expressed in terms of language specific constructions ("islands, specified-subject and other constraints on operations, Emonds's structure-preserving hypothesis, filters, etc.") and states that there were no attempts to relate any of this to "other biological systems" and that currently the "basic computational ingredients are considerably more abstract (locality, minimal search, basic recursion, etc.), and it becomes quite reasonable to seek principled explanation in terms that may apply well beyond language, as well as related properties in other systems" (Chomsky (2005)). There is no principled reason why locality should be compared with other *biological* phenomena but rather other *informational* phenomena. On the contrary, if locality were dependent on innate conditions, we might reasonably *not* look at other systems for examples of it. The trouble is, since locality was adopted from Rizzi (1990) by Chomsky and Lasnik (1993) into the MP, it has taken on the quality of a 'principle of UG', which remains a theoretical manifestation of the innatist hypothesis. As stated in Chomsky (2005), "The biolinguistic perspective regards the language faculty as an 'organ of the body,' along with other cognitive systems" and, furthermore, the "genetic endowment" is "the topic of Universal Grammar".

2.12. Our information storage/recall system is essentially conservative

If the function of the neocortex is to account for the elements of language so far described, it must be fundamentally 'conservative', prone to repetition rather than unconditioned variation. Anecdotally, young human children have essentially this character; they explore the world by imitation and are sometimes bewildered at apparently irregular occurrences (REF

reality of indices are implied here." By this he means to respect the 'condition of inclusiveness' (Chomsky, 1995b, 228) which suggests that nothing, including indices, should be added to the derivation during syntax.

for babies expecting continuity of shapes); they are especially attuned to regularities in their environment, linguistic and otherwise (Marcus (2004)).

2.12.1 Bruner and Postman, *On the Perception of Incongruity: A Paradigm* (1949)

Kuhn (1962) discusses the relevant psychological study Bruner and Postman (1949) at some length. Bruner and Postman asked subjects to identify various playing cards in short and controlled exposures. Although most of the playing cards are typical, some were specifically made so that the colours and suits did not match, e.g. a red ten of clubs or a black three of hearts. The experiments consisted of a series of single card exposures to individual subjects; as the series progressed, the exposure was gradually increased. Following each exposure, the experimenters asked the subjects what they had seen; if the subject correctly identified two cards in sequence the series was stopped.

The subjects were able to make identifications of most of the cards with only the shortest of exposures and with a small increase in exposures they were able to make identifications of all the cards. In the case of the normal cards, the identifications were mostly correct. Nonetheless, the abnormal cards were usually identified—without noticeable hesitation or confusion—as being normal, e.g. the black three of hearts might be identified as either the three of hearts or the three of spades. Seemingly without any awareness of the anomaly, the subjects categorized the abnormal cards according to known patterns of playing card design. After more (gradually increasing) exposure to the abnormal cards, the subjects began to feel something was wrong, stating something like, "That's the six of spades but there's something wrong with it—the black has a red border". Continued increase of exposure resulted in more confusion until the point where most subjects would, sometimes quite suddenly, make the correct identifications with confidence. Once having made the correct identifications of several cards, the subjects did not have problems with subsequent cards. Some subjects, however, were never able to make the adjustment to the abnormal suits, one stating, "I can't make the suit out, whatever it is. It didn't even look like a card that time. I don't know what colour it is now or whether it's a spade or a heart. I'm not even sure what a spade looks like. My God!"

For the purposes at hand, this is suggestive that the brain has a strong tendency to override anomalies and keep to previously established patterns, whether these might be typified as canonical word orderings, or archetypical word meanings or classes. In contrast to beliefs about objects in the world being triggers for innate knowledge (§2.8), this seems to indicate that, however these mental archetypes are developed, they are not limited to so-called 'innate knowledge'; it is hard to maintain that playing cards would be part of our innate knowledge. More reasonably, the fact that there are an uncontrolled number of exposures of words and associated (or non-associated) information to learners of language (etc.)

allows this information to be codified into what seems like innate archetypes. Nonetheless, my null hypothesis is that these patterns or referents are due to exposure and that they are signs that regularity is a necessary and evident trait of a learning system.

2.12.2 Stability in other systems

That child language is not highly irregular should only be astonishing if most systems were irregular. That most systems have high degrees of regularity—without the guiding hand of something analogous to biological UG—suggests that it is in the nature of systems to display regularity, e.g. cars would be hard to drive if their response to driver input were irregular or unpredictable; the functionality of our cardiovascular system likewise depends on regularity. Therefore the assumption that any system-acquiring-device (e.g. child learner) is prone more to repetition of language forms than to arbitrary or unconditioned variation seems correct.

I take it to be unsurprising, for example, if an adjective is commonly found before a noun in one language and after a noun in another language that these generalities should hold within the given language with minor and explicit exception[11].

2.13. The learner is pattern-oriented but not unique in this ability

That humans are essentially pattern-oriented is evident from tests done on human babies (e.g. Marcus (2000), Gomez and Gerken (1999)) but, as Marcus (2004) points out, this ability seems to be shared with other species such as the cotton-top tamarins (Hauser et al. (2002b)). This is not to say that these abilities are quantitatively the same but perhaps, to a reasonable degree, qualitatively the same. Clearly we outstrip all other species on the planet in our ability to find patterns in our environment but this is not an argument that we have a different means to find these patterns. The relevant distinction is that our biology is distinct (as any species necessarily is) but perhaps more distinct in a matter of degree than kind.

That we are generally adept at making and repeating patterns that are quite different than anything like what is found in our natural languages is suggested by our ability for complicated patterns in music as well as the various word games found around the world. Examples of these include pig latin, louchedem (Parisian, Lyonnaise butcher cant; similar to pig latin), rechtub klat (Australian butcher talk: words are spoken backwards), Cockney rhyming slang, jeringonza (Spanish wordplay game), verlan (French syllable inversion game). These are all essentially alternations of their respective natural languages but the relevant detail is that people seem able to alter their normal morphology *in real time* (albeit superficially and while maintaining phonological and syntactic norms). It shows that the ability for recursion and applying a *function* to alter lexical items systematically is not

[11] This in itself is not to explain why an adjective should be found adjacent to a noun in the first place but merely that the positioning, once established, can be expected to be consistent.

restricted to natural languages per se.

2.14. That spoken syntax has no consistent system of indices limits the system

There is no apparent manifest system of indices in the spoken information stream that aids the listener in making certain connections between words. In some cases certain lexical items are variably inflected with person, number and gender markers (collectively known as 'phi (ϕ) features') as well as morphological case; these seem to aid the listener to parse a given phrase or series of phrases, as in 14.

(14) a. I saw Sam and Angel the other day. I gave him a rock and her a flower.

b. I saw Sam_{fem} and Angel_{masc} the other day. I gave him_{masc} a rock and her_{fem} a flower.

In this example, the relevant people can be distinguished partly on the basis of the gender markings (if the genders are known to the listener); the word order could normally be an indicator but in this case it has been altered. What is not apparently available to the listener are the subscripts I have appended to the names and pronouns. Even on the hypothesis that indices (or their analogue) are available in the mind of the speaker, or even reconstructed in the mind of the listener, these are lost in speech unless marked by the manifest morphology (possibly by the ϕ or case markings). Because these markers are inconsistently available in the spoken syntax, it is a central point of this essay that their absence limits the possible forms of the syntax, as in (15).

(15) a. * How do you wonder why Catie degreased her bike chain?

b. How_j [do you wonder]$_i$ why_i [Catie degreased her bike chain?]$_j$

c. How [do you wonder] why [Catie degreased her bike chain?]

The reason we cannot understand (15a) is because there are no indices available to the listener as in (15b); nor are there arrows somehow available to the listener as in (15c).

A similar point is made in Chomsky (1995b) with the example in (16).

(16) how did John remember [*whether* Bill fixed the car *t*]

On the interpretation where *how* is associated with the trace and Bill's fixing of the car, (16) is illicit; but on the interpretation where *how* is associated with John's remembering, it is fine. That is, the phonological form has two syntactic/semantic interpretations that can be assigned to it. The distinction between Chomsky's analysis and mine is that in his movement is *blocked* by 'whether', whereas in my analysis, movement might be possible but *reconstruction to a canonical position* is confused by the presence of an intervening

like category, in this case 'whether'. This principle is closely related to the principle of Relativized Minimality of Rizzi (1990) in that the absence of indices is most detrimental when there is an intervening category as in (17).

(17) $\ldots\alpha\ldots\gamma\ \ldots\beta\ldots$

In the terminology of the day, Rizzi described the principle of Relativized Minimality as being: 'α cannot govern β if there is a closer potential governor γ for β'. My reformulation of this basic idea is: Where γ is an intervening and like category, α and β cannot share co-reference or be reconstructed to its canonical position. In the case of (15a) repeated here as (18a) this implies that 'how' cannot be reconstructed to its canonical position because of the intervening 'why', as in (18b).

(18) a. * How do you wonder why Catie degreased her bike chain?

b. $\text{How}_{\alpha}\ldots\text{why}_{\gamma}\ldots\_\_{}_{\beta}$?

2.15. Reconstruction & filters vs. attractors.

What is normally described as 'movement' in Chomskyan linguistics is assumed to be *argument displacement*. This can be considered to be constrained by the possibility of reconstructing the lexical item back to its canonical argument position (i.e. the argument position typical for the language in question). This is generally compatible with the 'copy theory of movement' (Chomsky (1995b); Fox (2002) etc.).

The distinction between these two views is a matter of the details. First of all, it is doubtful that there is a need for a discrete function to limit argument displacement of this nature. While theorists talk about the syntax being limited by such things as the head movement constraint (Travis (1984)), these apparent prohibitions do not have a clear mechanism. The definition of, say, Move Alpha as presented in, for example, Chomsky (1995b) provides little detail. The two most apparent logical possibilities seem to be as follows: i) the argument can (and would) move anywhere but it is constrained by such things as barriers or phases; ii) the argument could hypothetically move anywhere but needs a motivator such as features or Case to check. Since we are aspiring to the working of nature, I will make an analogy to biology. For a given function a hypothetical bacterium performs, the production of x is either limited by a negative regulator or by a positive regulator. This bacterium produces protein x only as long as z is plentiful (in positive or up-regulation); in the other scenario, it produces x until the supply of y reaches a given point and then production stops (in negative or down-regulation). The overall effect might appear to be the same but it is the regulating mechanism that is different. Both a decrease in Z and an increase in Y would cause the production of X to stop. This is a critical distinction if we want to build a detailed model.

My assumption is that arguments are free to move anywhere in a sentence (limited perhaps by working memory) but the syntax is constrained by the ability of the speaker/hearer to interpret the meaning of the phrase, i.e. the syntax is constrained by the semantics. If the phrase cannot be interpreted due to the failure of the parser to connect the various elements (e.g. because of the absence of indices, and the presence of intervening like-categories as in §2.14), this is a negative constraint.

2.16. The Circularity of Statistics

One of the central arguments that non-Generativists present and that Generativists shrug off is primarily that the Poverty of the Stimulus (PoS) is really not so impoverished as all that (cf. the "Special Issue" of *The Linguistic Review* "A Review of the 'Poverty of Stimulus Argument"' (Linguistic Review 19/1-2 (2002) with Pullum & Scholz 2002, Sampson 2002; vs. Fodor & Crowther 2002, Lasnik and Uriagereka (2002), Crain & Pietroski 2002 as representative examples).

An obvious problem with this debate is that the child has been exposed to the primary linguistic data for an uncontrolled period of time. Whether ±three years is sufficient or not to learn the basic patterns of a language is largely a matter of opinion and as such undecidable in the terms usually discussed. Furthermore it is not clear what period of exposure is exactly necessary because the recursive aspect of language or pattern making is developed at the same time as articulatory abilities and general memory.

2.17. Saussure's arbitrariness of the sign is not unique to humans

Clearly the arbitrariness of the sign, i.e. the ability to disassociate sign from meaning—*signifiant* from *signifié* of Saussure (1916)—is a central factor to human language. Nonetheless, it does not suppose that this ability is unique to humans. Pavlov's famous experiments with dogs suggests that dogs also share this ability. Any objection that the dogs were 'conditioned', and did not acquire this 'naturally' as children might, forgets that Pavlov certainly did no invasive reworking of the dogs' brains. If any physiological changes happened in the dogs' brains, these must have been due to a previously existing function of their brains.

Note too that this arbitrariness does not suggest that archetypes can in any special way connect with the two signs provided by the environment (cf. §2.7) and in fact if the two signs (signified and signifier) are truly arbitrary, they would have no gross identifiers consistent for a (possibly mechanistic) system to understand.

Note too that this is not to say language is a result of conditioning per se. Non-obvious forms of language (such as *wh*-movement) are partly as discussed in §2.14, although whether a *wh*-word can relocate or not is perhaps a result of conditioning[12]; the

[12]Note that current linguistic theory does not have a satisfying explanation for *wh*-movement. The theory of

sound (i.e. audible signature) of lexical items certainly is the result of such conditioning (see §2.9).

There is a tendency to assume certain abilities—that *are* biological endowments—as being particular to the human mind. As observed by Saussure (1916), the separation of sound from meaning is central to language use. Yet this ability is essentially the ability to recognize that a sound can be a representation of something other than the sound itself, something that even animals can do. Even if this ability is limited (a dog or a parrot might only ever recognize half a dozen 'words') the fact that they can do it, while in need of explanation, shows that this is not unique to humans. Birds can recognize birdsong and mating calls; presumably they are more interested in the other birds producing the sound than in the sound itself. Many animals use their hearing to recognize predators, prey or potential mates in some seemingly precise way. The critical point is that they have a means of reliably *identifying* the sound and relating it to another representation in their brains. If the human facility with language is more than this, it is not to say it is distinct in kind. It is very possible that it is only distinct in degree. If a bird somehow connects the representation of the mating call with the potential mate, it has made the first step toward understanding reference. That it may go no further may be due to a limit on its processing powers.

2.18. Spoken syntax is too impoverished and inconsistent to support a discrete computational system

Perhaps the most crucial postulate among all of these is the following. The problem of integrating a discrete system of language computation in the mind of the speaker alone has to date been a formidable one for generative theorists. Such an undertaking has arguably necessitated proposals such as (phonologically) *null operators* and *empty categories* such as *traces*[13] as well as *features* or sometimes (variably) overt or covert elements such as Case. These are argued to be necessary to guide and condition the grammar and according to many generative theorists, the grammar cannot be described systematically without these.

It must then be asked, if these elements are necessary for the brain to make sense out of syntax, *how is it possible that a listener, without access to the speaker's unspoken attractors (etc.), can make sense of spoken syntax?* Assume for a moment that the listener can make connections between, for example, *wh*-words and their canonical positions and permute these into an order satisfactory for the semantic component (i.e. LF) without the benefit of whatever features were necessary for the speaker to make sense of this same sentence.

features is highly stipulative and without empirical verification; it is currently the subject of much debate within the theory (cf. talk delivered by Norvin Richards at McGill REF). I speculate that *wh*-movement (fronting) is some kind of focal marker that has become codified in a number of languages; any quest for a more exact mechanism may be misguided.

[13] Or their basic equivalent, phonologically null copies.

If it is possible for the listener to either reorganize the sentence into something like LF without features (etc.) or interpret the sentence without such reorganization, *why is it that the speaker would need these features in the first place?*

In answer to this apparent paradox, this essay proposes that it is the impoverished nature of spoken syntax itself that constrains the forms the syntax can take. In other words, the system is constrained by listener comprehension. Even accepting the most highly detailed system of language yet proposed in generative grammar and allowing for principles barring 'look ahead' or the idea that the system can in any way 'learn', if a listener is unable to interpret a speaker's sentence, it is doubtful that the speaker would persist in speaking this way. For example, if the syntactic component of a speaker allowed *wh*-island violation but the listener—due to such things as a lack of overt indices—was unable to understand such a sentence, it is quite possible that the speaker would cease to speak these unintelligible sentences. My proposal is that we cannot know the inner workings of a grammar described in, for example, Chomsky (2000b) or Chomsky (2001) if the effects that in fact constrain the grammar are due to more pedestrian effects, the impoverishment of the spoken information stream.

2.19. Waiting for Godot...

Following from this impoverishment (§2.18), a final question must be posed. While it may be possible to describe certain aspects of language according to what seem to be rules, *the human capacity for language is quite possibly not rule based.* Therefore, a scientific inquiry that assumes that it is may very likely be trying to explain apparent breaches in this system according to null elements that 'complete the logic' when there is in fact no 'logic' to be completed. The brain may well not be interested in dotting its *i*'s and crossing its *t*'s in quite this way. If the essential neural function of the human brain is pattern recognition and, using this, human societies have developed some system that consists of patterns created and constrained by the mechanics of our articulatory and perceptual systems in order to communicate meaningful patterns to one another, we may have language that is only superficially concerned with rules. That is, the cortical tissue is not concerned with rules but the neural stimulation from the senses and consequently the possibility of recognizing patterns. If the patterns seem rule-like and help comprehension, so be it. If they are not needed but the overall 'message' is communicated, that is not a necessary barrier to the function of the pattern recognition system.

This is analogous to there being patterns in biological development that seem to be organized but are in fact only common to one another because of emergent properties of the problem at hand, e.g. sensitivity to light confers certain survival advantages and the ways in which an organ might be sensitive to light might have certain traits in common

with others without having been designed or following rules *per se*. Just as the human eye has evolved with a blind spot where the optic nerve attaches to the back of the retina and therefore has non-regular elements that must be compensated for, so too do systems of language have irregularities that must be compensated for (where often null elements are proposed in Chomskyan theory).

Our ability to *compute* language may in fact be fairly limited and we may be better described as sample, cut-up and playback devices (with some means of substitution of course). It is possible that the central component of our linguistic abilities is our memory and that we can only speak by modifying and compiling what we have heard before. Language may prove not to be motivated by compositional factors but be an emergent property developing from our talent for mimicry and recombination of associated sounds (and later used for more and more sophisticated communication).

These questions are perhaps not directly answerable at this stage. All we can hope to do is determine the forms of the grammar and account for them as best as we can, with the least supposition as to the underlying biology. When we do get to the point where we can better understand what is happening on the biological level, questions as to the *discrete-continuous* interface will likely be critical. We need to understand how it is that we can perceive spoken strings of sound and correlate these with whatever representations and previous representations we have of sounds in our brains.

From a biological point of view, any applications of rules may be irrelevant for purposes of the brain functions that pertain to language and any rules that are relevant to comprehension of language, while conditioning aspects of the output of the language (i.e. the character of a particular language or phenomenon of a language such as *wh*-word interpretation) may be extra-phenomenal, more an emergent property of the system than an integral part of neocortical language functions. To address language in a scientific manner we might investigate the possibility that we are not as sophisticated as we might like to believe, that the brain has a means of observing, reproducing and, to an extent, permuting information from its environment but that it does not engage in computation in a sense even remotely like a digital computer. This is not to underestimate human cognitive abilities but to attempt to correctly estimate without consideration to our lay assumptions about these abilities. If we want to limit talk about 'mind'—keeping it as a clear emergent property of the brain—we need to examine what it is that the brain is, how it is composed and how the functions we witness might be accounted for under a less specialized model.

Certain of the issues addressed in this essay cannot be resolved—or perhaps even known—until we have made more progress in neurological sciences. While I make some reference to generalized memory, my arguments do not depend on exactly what this may

turn out to be[14]. Nonetheless I suggest that it is a mistake to allow these issues to be confused by rhetoric arguing the 'likelihood' or otherwise of certain phenomena (such as language acquisition or the generation and understanding involving inflection and other licensing nodes for lexical items) requiring one to accept the idea of a discrete 'language organ' in the brain. For the purposes of this discussion, I suggest that until we have confirmation for such hypotheses, we might evaluate the restrictions on language itself so that we expect the least distinction between us and our evolutionary ancestors that still describes the data.

As such, our ability to generalize and substitute linguistic forms for others is an example of our general abilities in other domains such as visual or auditory recognition of permuted forms (e.g. facial or voice recognition) and as such a part of general higher cognitive functions that are unique to our species, possibly due more to reasons of quantitative than qualitative differences in brain structures. These are, nonetheless, not issues that we should attempt to resolve through rhetoric or design polemics as though we are engineers debating personal issues of design economy and likelihood or not of learnability. Because we cannot understand these in terms of even the vaguest representation of neural organization, I will not propose hypothetical structures but will examine other interfering elements that restrain the system.

In all of this we must keep in mind that there are inevitable difficulties in defining how language might operate simply because we are studying ourselves and studying a very personal aspect of ourselves, i.e. not something dispassionate like the liver or the quadriceps, but the brain and how language might operate in brain. Quite likely, even once we have an adequate theory of how neurons work together to produce the effects we are now familiar with, any theory of 'mind' will still be prey to lay philosophical controversy. That is, even once we have a good understanding of the brain, our personal theories of mind will interfere.

[14]For an interesting proposal, see Hawkins (2004)

Chapter 3

The Logic of Lexical Categories

3.1. Understanding, identifying and keeping track of lexical categories and lexemes

Before we can discuss syntax in any meaningful way we must quickly take a look at what syntax seems to be comprised of, words or morphemes, i.e. lexemes. Syntax seems to be the process of how these lexemes might be indicated, modified and ordered. Nonetheless, while we can list seemingly authoritative descriptions of these, we still have a very poor understanding of what they really mean to the brain. As Bloom (1994) states, there is little consensus of what it means to 'know a word' (see Carey (1982), Lakoff (1987), Premack (1990) for relevant discussion). According to Bloom, child word acquisition is one of the biggest 'mysteries' in the field of child development. According to Carey (1978), the vocabulary of a child grows at approximately nine words per day from ages of one and a half to six years. As Bloom points out, acquisition theories cannot be considered complete until we understand what it is that children are acquiring. My suspicion is that they mean a lot *less* than we usually assume; by this I mean that I suspect that the meaning of lexemes is more an issue of some kind of crude 'triangulation' of past usage (and as such the 'lexicon' is actually part of general memory) than what might be assumed by advocates of a 'discrete lexicon'. (See comments in Wittgenstein (1958) and §2.8). It is very possible that our logical analysis of word meaning is not the correct approach, at least not if we want to understand how the brain is processing this information.

I assume here that the matter of lexical representations may be a truly biological question (i.e. to be reserved for neuro-biologists). This may well derive from the biology underlying the comprehension of the Saussure (1916)'s arbitrariness of the sign and that they understand that sounds can represent objects in the world. What seems to be the case is that the arbitrariness of the sign is an ability we share with other animals; although we can remember and recombine these tokens to a much greater extent, it seems a mistake to assume that this is something particularly human. If a horse understands a verbal command by tone, sound or intonation this is an example of their understanding the arbitrariness of

the sign—even if they only understand it for one command. That is, our abilities are clearly *quantitatively* superior but not, at least in this regard, provably *qualitatively* different.

What can perhaps be answered from a linguistic point of view is whether or not lexical categories, such as verbs, nouns and adjectives, are preternatural classes with arbitrary bases. I argue that they are essentially non-arbitrary and that these classes have no special powers (e.g. ϕ-feature and tense/agreement checking). I assume this simply because the classes themselves, while perhaps codified by history, seem to be a non-arbitrary answer to the problem at hand, putting the arbitrariness of the sign to work to refer to the world and communicate with one another. This is the subject of the following section.

One reason for this assumption is that, as discussed in §2.7, an operational system that needs to process data (etc.) according to kind necessarily needs to be able to recognize a marker indicating the 'kind' in question. As such, if lexemes are processed other than by relational position with other lexemes (i.e. if there were a discrete computational function of grammar), it would necessarily depend on some kind of identifier or signature on these classes. The absence of any such identifier, coupled with the non-arbitrary nature of these classes (the focus of this chapter), suggests that the grammar must be organized in some way so as not to need discrete reference to the lexical category of a lexical item.

3.2. Binary features and lexical category

Since Chomsky (1970) Generative grammar has largely assumed the binary feature system of $\pm$ N and $\pm$ V in the following configuration:

(19) a. +N, -V = noun

b. -N, +V = verb

c. +N, +V = adjective

d. -N, -V = adposition (post- or preposition)

As Baker (2003) notes, "this theory is widely recognized to have almost no content in practice" and largely do not fit into any other parts of the theory. He further notes that the features system in (19) is "more or less arbitrary" and cites Stuurman (1985) (ch. 4)and Déchaine (1993) (§2.2) as authors who argue that it is possible to find syntactic evidence "in favour of any logically possible claim that two particular lexical categories constitute a natural class". He also notes that, Stuurman concludes that "the idea of decomposing syntactic categories into complexes of features is bankrupt" (Baker, 2003).

He notes that in the linguistic literature, it is generally assumed "that the category of verb is universal" and that criteria of "category-neutralization that have been suggested generally work in favour of the verb". Which is to say that in the cases of the "many

languages [that] are said to have no adjective-verb distinction" (e.g. Mohawk, Choctaw) the theorist takes the common category to be a verb, not an adjective. In the case of languages in the Wakashan and Salish families (found in British Columbia and Washington State) some have claimed there is no noun-verb distinction. Here Baker writes, linguists are "shier" about deciding between the two categories.

3.3. Baker's Reference-Predication Constraint

Nonetheless Baker himself wants to maintain the idea of lexical categories. He summarizes lexical categories (as opposed to functional categories) into nouns, verbs and adjectives and proposes the "Reference-Predication Constraint (RPC)" and considers this a "cornerstone" of his theory of lexical categories (Baker (2003); Baker (2004)). The RPC is as in (20).

(20) The Reference-Predication Constraint (RPC):
No syntactic node can have both a specifier and a referential index.

Baker's argument in favour of the RPC is roughly as follows. He notes that, while it is generally possible to derive verbs from predicate adjectives, deriving verbs from predicate nouns is much less consistently productive. Baker (2004) uses the example in (21) from Hale and Keyser (1993).

(21) a. The screen is clear
b. The screen cleared
c. Chris cleared the screen

(22) a. John is a man
b. * John manned (= 'John became a man')
c. * The war manned John (= 'The war made John into a man')

Baker claims that the RPC explains this difference because, although an adjective can be made to act as a verb "by the simple, monotonic process of endowing it with a theta-role that it can assign to its specifier", this is not possible with a noun. The reason he gives is that, according to his own RPC, if one assigns a noun a thematic role (theta-role), there is a violation of the RPC. In Baker's generalization, the only way a noun is able to act as a verb (i.e. be endowed with a thematic role it can assign to its specifier) is if it does not also have a referential index. The only way therefore for a noun to be lexically well formed is "if the nouns inherent referential index is also suppressed in some way". Otherwise, Baker claims, it is not possible for a given lexical item to be both a verb and a noun. Baker states that this

is meant to reflect the apparent fact that a given category cannot at the same time refer and be a predicate (citeBaker:2004)[1].

3.4. Towards a non-arbitrary definition of Nouns and Verbs

Despite the seeming correctness of Baker's analysis, we are still left with an essentially arbitrary system. Which is to say that, while Baker correctly determines a generalization, he does so by assuming two traits to verbal and nominal categories that are not apparent in the spoken syntax. That is, as much theoretical underpinning as specifiers and referential indices might have, we are given no principled reason why in fact a verb should have a specifier in the first place nor, for that matter, why a noun necessarily bears a referential index. While it is entirely possible that verbs and nouns are arbitrary manifestations of how we are 'wired', i.e. how we are innately disposed to see the world, such an assumption should not be the null hypothesis; it is desirable, if possible, if we can derive nouns, verbs and adjectives from more basic principles (see §2.2).

We can say with some truth that objects (concrete or abstract) in the world are often described by or subject to *states*. Active ones we might call events and static ones we might refer to simply as states. But the problem here is that while in some languages these nominally static states might be described by stative verbs while in others they are described by adjectives. An example of how the stative-verb/adjective distinction blurs can be seen in the following example:

(23) a. ka- hútsi
NsS- black
'it is black'

b. t- a'- ka- yá't- ∧ '-ne
CIS- FACT- NsN- body- fall- PUNC
'it (e.g. a cat) fell'

(24) a. ka- rák- ∧
NsS- white- STAT
'it is white'

b. t- yo- ya't- ∧'- ∧
CIS- NsO- body- fall- STAT
'it has fallen'

3.5. Nouns: Things that refer

Baker (2003) tries to define nouns by isolating how they are unique compared with other

[1]This division is tested by the nature of gerunds, most explicitly and directly by Malouf (2000) and Hudson (2003). Baker states that "The standard generative insight about gerunds is that they have the external syntax of a nominal projection and the internal syntax of a verbal projection", as discussed in detail in Baker (2004).

categories, shown in (25).

(25) Baker's defining characteristics of nouns

a. Semantic version: nouns and only nouns have *criteria of identity*, whereby they can serve as standards of sameness.

b. Syntactic version: X is a noun if and only if X is a lexical category and X bears a *referential index*, expressed as an ordered pair of integers.

Baker explains that the semantic version (25a) is from Geach (1962) and Gupta (1980) via Larson and Segal (1995). Informally, Baker describes these criteria as, "only common nouns have a component of meaning that makes it legitimate to ask whether some X is the same (whatever) as Y" (96) and specifies that, contra common conceptions (e.g. in Croft (1991)) that nouns are *inherently* concerned with the business of reference. Rather they, as discussed in Hopper and Thompson (1984), indicate "discourse manipulable participants" which is to say they are suitable for tracking a referent—as described in Wiggins (1980) (chapter one), indicating the same "entity" again and again.

Baker states that his categorization indicating nouns as reference trackers—following Geach (1962) and Gupta (1980)—within discourse more accurately captures certain facts about quantifiable expressions and predicate nominals than a theory that considers reference as part of the fundamental nature of nouns. Clear, in example (26), there is no act of reference implicit in the respective nouns.

(26) a. No [$_{NP}$ letter(s)] arrived today

b. No [$_{NP}$ wine] is served during Lent

Nonetheless, Baker considers the "special referential powers" of nouns to be a clear corollary of the inherent traits in (25). On this point, it is worth pointing out that this choice between theoretical approaches is not trivial. The two general options seem to be i) nouns are more inherently reference trackers within discourse with a side effect of referring to objects in the world or, ii) nouns are referrers to things whether or not they are in fact in the world, i.e. whether or not they refer to negative instances of things. Leaving aside the question of quantifiers, there are other serious problems with the noun-as-referrer approach. This is to say, from a semantic point of view both negative (as in (26)) and fictitious instances of things seem to indicate that nouns are not at base referrers; for reasons including logical impossibility and storytelling, some nouns, as in (27), are not part of our world and cannot be referrers.

(27) Referrers to logically impossible referents[2]

[2]In standard semantic theory this sentence is a problem because composition of the set of all objects former

a. The present king of France is bald

b. An American king visited Utrecht

(28) Referrers to fictitious worlds

a. Alice played Pac-man with the White Queen

From a purely analytical point of view—i.e. upon analysis of *adult* grammar—Baker's formulation seems to be an attractive generalization. That is to say, some nouns, such as certain quantified expressions, *never* refer to anything in the real or even fictional world. Nonetheless, from a point of view of how language is actually used, particularly by children in their one word stage, it is not very satisfactory. While it is perhaps a more elegant generalization, elegance must sometimes give way to accuracy. Children seem to understand the basics associated with the generalization in Geach (1962) and Gupta (1980), that nouns constitute a system of regularity or *sameness* with regards to an object in the world and the associated sound or sign[3].

3.5.1 The Child's Environment & Reference

If nouns were primarily concerned with reference tracking inside discourse—a logically, and even empirically, tenable hypothesis—one might find that children are less sensitive to nouns *as referrers*. Nonetheless, despite the occasional claim that the child's input is 'chaos', the data that children actually end up learning is actually fairly orderedor at least constrained by immediate context. That this is the case is suggested by the fact that Steven Pinker, one of the most prominent advocates of the 'language instinct' hypothesis explains that *mere instances of language are not sufficient for language acquisition*. If all that it takes for children to learn language was exposure to language then they should be able to learn language from the relatively non-chaotic environments of television or radio. Pinker writes:

> Children do not hear sentences in isolation, but in a context. No child has learned language from the radio; indeed, children rarely, if ever, learn language from television. (Pinker, 1995, 156-157)

Both radio and television are rich in linguistic data but are not constrained contextually clear environments from which to get a toe-hold on language—or possibly the idea of

---

with the set of all presidents results in an empty set; according to this interpretation, the verb cannot even begin to compose with the subject argument; the problems with these sentences sometimes are hard to grasp, perhaps due to the almost routine casualness with which we refer to things that cannot be referred to; perhaps because we are very willing to imagine fictitious worlds (again, this makes any notion of a necessary level of LF questionable—we seem not to have a problem understanding things that apparently make no sense)

[3]Ultimately this dichotomy may be without substance. I speculate that the relation may be something along the lines of Saussure's arbitrariness of the sign, where the sound refers to the representation and the representation refers to the world or possibly another (fictional or impossible or otherwise imaginable) referent.

language itself—we do not know if children must rediscover this each time or not[4]

This is to say, the environment must in fact be highly constrained even if no naïve ideas about 'motherese' or teaching possibilities are entertained. One might speculate that a language organ hypothesis should expect a healthy infant mind to recognize and begin to parse human language regardless of the medium[5].

Specifically, as regards nouns and reference, it is worth noting that the one word stage in child language is not a random catalogue of words. This is suggested by the fact that their selection for objects contrasts with their selection for words for event and for people:

> "Taken as a group, the object words in the single-word period form a broad semantic class which contrasts with other semantic classes emerging at the same time. That is, the pattern of usage of object words contrasts with that of words for events ... words for persons ... words for temporary states, greetings, and negation, and so on" (Huttenlocher and Smiley (1987) quoted in Bloom (1994)).

This seems to indicate a strong ability to categorize but, notably, an ability that does not necessarily correspond with the categories adults or linguists consider relevant. For example, greetings are not a significant class for adults or linguists but quite relevant in a child's world.

Furthermore, supporting the study mentioned above (Pinker, 1995, 156-157), Fisher et al. (1994) discusses studies conducted by Gillette and Gleitman ("Effects of situational cues on the identification of nouns and verbs") where the adults were asked to watch a video of five to ten minutes, with the sound muted, of mothers interacting with their children (<2 years). The instruction to the subjects was as follows: a beep is sounded at certain times when the mother speaks; the subject is asked to guess what the noun represented by the beep is. In the case of common mother/child nouns, the subjects guessed nearly always correctly and usually it only takes one 'beeped' scene viewing to correctly guess the relevant noun. Fisher et al. (1994) takes this to mean two things. First of all, this suggests that the input situation of mothers with very young children tends to have mothers speaking of objects that are the focus of the discourse, perhaps being manipulated by either the mother or the

[4]Would not a wolf-child create a personal language if language were a biological imperative? (This is not a necessary supposition but worth considering, cf. babies otherwise exploring their environment and body.) Pinker and Jackendoff (2005) discuss this case, (perhaps) inadvertently weighing in on the side of language-as-tool-for-communication: "Turning to cases in which languages are invented, we find that Nicaraguan Sign Language, for example, arose in the context of a community seeking communication (Senghas & Coppola, 2001). Similarly, isolated deaf children who create home signs do so in the context of communication with others. We are unaware of cases in which deaf individuals develop a complex vocabulary and grammar just to talk to themselves. And without exception, other linguistic isolates do not develop speech at all." (6)

[5]In this I am suggesting that the child needs some contextualization as a means of bootstrapping. Of course the environment is not too specific. In a constrained environment a parent is very much less likely to call out "Honey I am opening the door" as opposed to "Honey I am home".

child (making recovery from context a much easier matter). Fisher et al. (1994) cites Bruner (1975) and Slobin (1975) for similar studies. The second conclusion they draw regards the interpretation of the information given by the situation; the adults were able to accurately assess the "level of specificity at which the speaker is making reference—elephants rather than animals or puppets—despite the fact that all of these interpretations fit the observed scenes equally well" (335).

Whether adult deduction levels are the same as those of children is another question but perhaps the act of induction in both cases is similar enough for the purpose of the study. The critical question comes in regarding the basic principle of the *arbitrariness of the sign* (Saussure (1916)). That this ability for deduction is itself not a uniquely human ability (as discussed in §1.1), is shown by the fact that other mammals can associate arbitrarily related *signifiant* and *signifié*, such as the dogs in Pavlov's famous experiments.

Even if nouns are primarily referential, once pronouns are introduced and children acquire the ability to use more complex phrases, including verbs, the idea of reference tracking becomes more relevant. This idea is explored in §5.1.

3.6. Quantifying NPs, Logical Form, Quantifier Raising & Distributivity

In order to establish the non-arbitrary nature of all lexical categories, this section briefly investigates the subject of quantifying noun phrases, aka. quantifiers or QNPs. The special nature of quantifiers, I suggest, is not as mysterious as usually assumed and furthermore, the means of describing the interactions of QNPs is largely a holdover from the philosophical traditions from which they originated and do not take into consideration basic principles of computation.

The principle of Logical Form has been, since at least Chomsky (1981), considered motivation for a good deal of covert movement. Such covert movement is taken to imply both that a lot of unspoken reorganization is going on in language (motivating, in part, theories of agreement when there is no overt agreement). More significantly, it been taken to be a unifying factor in human language. As much as languages vary on the 'surface' (spoken syntax), if we accept the T-model of Chomsky (1995b), it is hypothesized that word order in languages attain the same basic configuration at the level of Logical Form (LF) and vary only in whether certain components of the syntax move overtly or covertly; if they are 'spelled-out' (spoken) before they move (e.g. Chinese), they appear fairly similar to base or canonical positions at the surface, whereas if they are spelled-out after they move (e.g. Bulgarian), they are considered to more accurately reflect the order supposed to be needed by the semantic component. This theory provides an appealing explanation for cross-linguistic variation but, while perhaps consistent with a discrete computational language component, suffers from lack of direct evidence. Central to the evidence accumulated, is the theory of

quantifier interpretation and with it, distributivity.

I use the term 'distributivity' after Beghelli and Stowell (1997) as a means of pinpointing the phenomena in question. By this I mean what is normally discussed in terms of scope; 'every' is normally thought to distribute over 'some' if it scopes over it via quantifier raising (QR). Which is to say, distributivity is not considered a primitive of quantifier theory but a central empirical test that is standardly thought to characterize the relationship between the universal $\forall$ and the existential $\exists$ quantifiers when they are read in the configuration $\forall > \exists$ and not when read in the configuration $\exists > \forall$. Nonetheless, I attempt to promote this 'test' to the status of 'the thing itself' and suggest that QR and, by consequence, LF are theoretically bi-products—or singular (but possible) interpretations—of the data from this test.

In order for the relation between $\forall$ and $\exists$ to be logically significant the first must distribute over the second. What is not clear in the analysis of natural language is whether this must be achieved by a process similar to QR, where the quantifiers that occur in an inverse relation, reorganize through covert movement or if there are other ways the language faculty might interpret this relationship.

The standard rule of scope is as in (29).

(29) α scopes over β iff α c-commands β

Given standard assumptions, for an object quantified noun phrase (QNP) to scope over the subject and be interpreted, it must move to a c-commanding position relative to the subject, as in (30).

(30) Everybody somebody loves t $(\forall > \exists)$

Likewise, in order to get surface scope the subject then could move over the already raised object as in (31). Alternately no quantifier movement need have taken place and the operators stay *in situ*.

(31) Somebody everybody t loves t $(\exists > \forall)$

3.6.1 The Order of Operations

As stated in Heim and Kratzer (1998), "On the relational theory of quantification, quantifiers denote relations between sets. For instance, 'every' denotes the subset relation, and

'some' denotes the relation of non-disjointness". If in fact we are discussing relations between sets it then becomes relevant to understand better the nature of the sets. A very basic distinction between sets is what it is that orders the computation of a set and whether such elements as negation might play a role in this (more than just being a blocking agent to movement). These factors will necessarily play a role in regards to quantification.

If sets are mapped according to function application, it is not clear why computations would necessarily operate in one order over the other. I suggest that, in keeping with principles of computational economy, once the lexical items have been identified as 'logical operators'[6], it is possible that the brain chooses intonational or pragmatic cues to decide what order to process these operators in without necessarily reordering them through covert movement. That is, once we have a theory that presumes there are movement operations, presuming QR as a process of covert movement to reorganize the logical operators may not be considered computationally costly. Nonetheless, principles of computational effort in other domains suggest that readings a string of operational commands is as simple backwards as forwards, upside down as right-side-up[7]; operations that require that certain operators are read before others may therefore be read as retrograde strings without undue effort. In mathmatics for example, computation of a string such as in (32) or (33) may be sensitive to the direction of computation but it would be unwarranted to presume that we can only compute in one direction or the other[8].

(32) a. i. $\Rightarrow (4 \times 2) + 6 = 14 \Rightarrow$

ii. $\Leftarrow 14 = 6 + (2 \times 4) \Leftarrow$

b. i. $\Rightarrow 6 + (2 \times 4) = 14 \Rightarrow$

ii. $\Leftarrow 14 = (4 \times 2) + 6 \Leftarrow$

(33) a. i. $\Rightarrow (4 \div 2) - 6 = -4 \Rightarrow$

ii. $\Leftarrow -4 = 6 - (2 \div 4) \Leftarrow$

b. i. $\Rightarrow 6 - (2 \div 4) = 5.5 \Rightarrow$

ii. $\Leftarrow 5.5 = (4 \div 2) - 6 \Leftarrow$

That a parser would be sensitive to the order of operations follows from interpretation of the logical operators just as it does in mathmatical operations. That is to say, it is unneces-

[6]Whatever this might mean. I maintain the terminology for the present discussion despite my deep scepticism of the idea that 'logical operators' have a physiological reality. Which is to say, just as the child can 'learn' that verbs take noun arguments (perhaps, in a sense, simply a form of connecting one lexical 'idea' with another), so too do they observe that quantifiers imply more connections of a slightly different sort when in certain configurations.

[7]For example, it is well known that information from the eyes enters the brain 'wrong-way around' and is corrected, presumably with little computational effort.

[8]Alternations of, in this case, bracketing will also change the result (e.g. $4 \times (2 + 6) = 32$); whether there is useful analogy to distributive or non-distributive readings is another question.

sary to presume that the type shifting typified as QR necessarily involves covert movement of these operators. Perhaps, as in mathematics (and such things as the directionality of written language[9]), the matter of which direction a parser operates is perhaps partly a matter of agreement on certain linguistic 'conventions' (see discussion on page 30).

(34) a. DISTRIBUTIVE

⇒ Everyone loves someone ⇒
⇒ ∀ → ∃ ⇒

'For every person$_x$, there exists some person$_y$ such that x loves y'

b. NON-DISTRIBUTIVE

⇐ Everyone loves someone ⇐
⇐ ∀ ← ∃ ⇐

'There exists some person$_x$ such that every person$_y$ loves x'

(35) a. NON-DISTRIBUTIVE

⇒ Someone loves everyone ⇒
⇒ ∃ → ∀ ⇒

'There exists some person$_x$ such that x loves every person$_y$'

b. DISTRIBUTIVE

⇐ Someone loves everyone ⇐
⇐ ∃ ← ∀ ⇐

'For every person$_x$, there is some person$_y$ such that y loves x'

Finally, if we count the correct interpretation by a listener as a necessary part of the question under examination, that the listener can decode intended quantifier order *by deciding which logical operator should be read first and not by somehow detecting covert movement*, we have to presume a mechanism that allows operators to be read inversely without necessarily being displaced. If an operation involving inverse computation exists in a listener, it might be expected that a speaker would also be able to read logical operators in a syntactic string inversely.

Such retrograde functions may be operational in other areas of the grammar where covert movement is presumed to apply; compare (36)

(36) Because he was worried about his mother's operation, John didn't get much sleep
← he$_i$ ← ← John$_i$ ←

'John was such that, because he was worried about his mother's operation, he didn't get much sleep...'

[9]Chinese is an example of the arbitrariness of such conventions. Just as it is apparently ethnocentrically arbitrary to say left-to-right is prior for reading, so is it perhaps more due to the routine aspect of speaking in a time-based context that biases us towards thinking of that necessarily being our means of processing operations.

The question to be addressed with regards to this generalized proposal is how the first linear operator in either the quantified phrase or the 'indexed' pronoun-R-expression phrase is regarded until both operators are stated; perhaps given certain configurations of 'missing information' a sentence is parsed once, then parsed a second time to establish the required reference relations[10].

If we assume the interpretation takes constituent structure seriously, sentences such as in (37) are not a problem; i.e. reading a given constituent for the variables inside it and if there is an embedded constituent containing its own relevant variables that do not interact (or interrupt) then we do not anticipate a problem.

(37) a. Most accidents that nobody reported were minor

b. one apple in every barrel is rotten

c. Sue read every book John did

d. Someone from every city despises it

This approach would explain why intonational focus plays such a disruptive role in quantifier interpretation: giving a quantifier sentential stress is thereby tantamount to saying, *Interpret me first!* The LF component on this view would not be a discrete system or component but simply a way of modelling the final interpretation of a very adaptive and flexible parser that is sentitive to pragmatics and prosidy. The example in (38), with sentential stress on the object normally results in a distributive, inverse scope reading (Herburger (2000)).

(38) Some girl loves EVERY GUY in the dorm

On the view that the parser can compute backwards and forwards, sentential stress may simply be an indicator that computation should start with 'every' and proceed to the next logical operator in the domain.

The point I wish to make here is simply that the data does not necessarily imply that a movement operation takes place in situations involving inverse scope. For this reason I will not consider the data with the idea that QR is the only way to model the relevant type shift. Likewise, in this essay (as implied in §2.15), I do not examine intervention effects (on a move-α model) that follow from a theory that quantifiers either move or do not move to establish their scope (Fox (2000) and references within).

The reasons for this are as follows: the speaker only ever receives data for interpretation stripped of any rich specification such as whether covert movement will apply, whether

[10]Perhaps experimental evidence on processing time could shed light on this question.

a NP is situated on a node giving it special properties or, in a morphologically impoverished language and if the configuration of XPs present is similar to an underlying form (s_1), whether the string of sounds are a derived form (s_2) and therefore distinct in certain non-trivial ways. This is to suggest that if the listener cannot distinguish s_1 from s_2, the interpretative component does not need access to internal processes and will compute certain discrepancies such as quantifiers that are expressed other than in the logically intended order. A reason for believing that the listener might not need to know the level of derivation but instead *intakes* a string of sound and then computes how to read the string, is such reasoning as in Abney (1991), where he discusses the idea that a human sentential parser may proceed by taking a 'chunk' at a time and interpreting that.

If we understand quantifiers not as being *special* properties, but as being natural results the logic of natural language, we can understand better our writing conventions and the place of quantifiers in natural language. In a sentence such as "$\exists$ girl$_x$ $\forall$ boy$_y$" we may choose a default interpretation involving a singular girl. That is to say, there is no reason we should suppose there is more than one girl and therefore a distributive or list-pair reading is not expected if the $\exists$ operator precedes a $\forall$ operator. On the other hand, in a sentence such as "$\exists$ bite$_x$ $\forall$ boy$_y$" we do expect a plurality of bites because we know that once bitten, the bite cannot apply again to another boy. Therefore if pragmatics or other factors tell us to call into existence more than one thing we are able to think about distributivity or list-pair readings in ways other than determined by first order logic (a human invention after all). As a thought experiment, imagine the statement "$\exists$ girl$_x$ $\forall$ boy$_y$" where it is specified that there is a multitude of girls. Here a list-pair reading is possible, if somewhat odd suggesting that the distributive nature of a $\forall > \exists$ order is not due to the special nature of the quantifiers but due to the logical meaning of attributed to them[11].

3.6.2 Possible interrupters to bi-directional parsing

In (39), this failure of distributivity can be contrasted with the behaviour of the same logical operators outside of the DP, as in (40) where inverse distributivity is possible.

(39) a. [Someone's love of everyone] in my family doesn't surprise me ($\exists > \forall$; $*\forall > \exists$)

b. [Everyone's love of someone] in my family doesn't surprise me ($\exists > \forall$; $\forall > \exists$)

(40) a. That someone loves everyone in my family doesn't surprise me ($\exists > \forall$; $\forall > \exists$)

b. That everyone loves someone in my family doesn't surprise me ($\forall > \exists$; $\exists > \forall$)

[11] Though not unlike a lambda calculus where we know what to expect.

While it is possible to describe this data in other terms[12] so as to maintain a theory of QR and LF, it is also possible to view this as a case where there are certain 'interrupters' to bi-directional parsing. That is, while QR is a means of modelling type shifting, it is entirely possible to analyse the data in in terms of the type shifting itself and without presuming that QR is a precondition for this sort of type shifting.

3.7. Verbs: Animating Referrers

If children and adults are concerned with referring to, and describing, their world, it is reasonable that they are concerned with making reference the state or description of it. In languages such as English the division of *states* into something like active states, i.e. *events*, and static states, represented by stative verbs and adjectives. Nonetheless it is difficult to imagine expressing a state without a referent, whether to a material or abstract object

As discussed briefly in §2.7, Baker (2003) defines verbs in the following way. They are described in a somewhat semantic way, according to how they relate to other lexical categories, as in (41a); Baker alternately defines them structurally, as in (41b).

(41) a. verbs "are inherently unsaturated expressions that hold of something else, and thus the nucleus around which sentences are typically built"

b. X is a verb if and only if X is a lexical category and X has a specifier

(23).

Yet in both of these definitions, we have a very particular function (or structure) associated with a given phonetic shape, e.g. 'walk' [wôk] (that allowably can be generalized to a particular phonological form). Even allowing that this phonological form can be identified as a member of a category, i.e. verb, (but cf. 2.7), is this reasonably the most basic level of description? It is of dubious merit to ignore the fact that, as verbs are used in language, they have no means of referring. If language is concerned with communicating and if communicating is essentially a system of referring then verbs, as we use them, are in need of elements that do refer, i.e. nouns. This is to say, we cannot easily imagine or define a verb without assigning it at least a stand-in referring item. If we want to define 'jump', we need to perhaps imagine a boy or a frog or *something* jumping. For this reason, I here define verbs as a class of words concerned with shedding light on the state (active or passive, present or past, etc.) of referrers and as such dependent on nouns for any sense.

If, as suggested in (41), Baker's informal description of verbs ("inherently unsaturated expressions that hold of something else, and thus the nucleus around which sentences are

[12] As, for example, Richard Larson (pc. via Sabine Iatridou) does. He assesses this data in terms of his analysis of Determiner Phrases (DP), and that "pronominal genitives are derived by the equivalent of dative shift ... hence the scope freezing in nominals becomes an instance of the same thing one finds with DOCs."

typically built" (Baker, 2003, 23)) can be made into a perhaps cruder but more concrete expression, along the lines of (42).

(42) a. verbs are inherently descriptions of events or states, communicating temporal and spacial relations

b. in order to describe, one must describe *something*

c. verbs do not themselves containing a means of referring

d. as such verbs are in need of referrers, i.e. nouns or their analogue

e. verbs therefore must form a clear relation with referrers such as nouns, whether by local positioning or by unambiguous linking identifiers (see §2.11)

This generalization is corroborated by statements by Fisher et al. (1994) arguing that "verb learning implicates a sentence-to-world pairing procedure" (comparable to nouns being a "word-to-world pairing procedure"). In general locality to the noun (or the reconstruction position of the noun) is important if there is no discrete means of connecting a noun-verb or noun-verb-noun grouping.

3.8. We do not want to be botanists

What this work by Baker and others suggests, is that, for the very reason that determining a 'natural class' of word categories is so problematic, we might consider giving up the idea that there should be a definite natural class. The data is so conflicted that, if Baker only narrowly teases out a difference between verbs and adjectives (both in a way modifiers of referents) then perhaps the brain is not so rigid in its interpretative assumptions and in fact is not processing these classes in any kind of formal and discrete lexicon.

As interesting side-note, Baker points out that we do not out of conceptual necessity need nouns in the sense we know them. He agrees that languages seem to need referrers but this role he claims could be handled by "functional categories that bear referential indices, such as pronouns and determiners" (170). His example for a nounless language might include something like 'child' being stated as an adjective 'the childlike one' or as a verb 'she who childs'[13].

We could categorize and organize like botanists, disagreeing and agreeing in varying degrees. The real question seems to have escaped the debate: does the brain pay any attention to these classes, i.e. do they have any reality in any important sense? If there are merely

[13]Taking this idea one step further, one might take the idea that nouns are reference trackers and consider the ambiguity between childlike+one and child. One might choose to describe most nouns to be an 'it' with a built in description just as most verbs might be considered as a go-between, mediating the referents and the description tagged to them. We could envision a language with nothing but a word such as 'it', a plural marker and a list of adjectives, describing time and other details affecting this referent word.

manifestations of logic and perhaps a little history or culture then they are not interesting for the investigation at hand.

Having examined the various qualities of the lexical categories and determined that there is nothing particularly arbitrary or unexpected about them, the next step it to attempt to understand how and why they fit together as they do.

Chapter 4

Composition, Adjunction & Theories of Acquisition

On the assumption that there is no innate knowledge of what a sentence should be, how do children learn sentences? The question turns on whether we consider the syntactic forms found in human languages are essentially *arbitrary* or *non-arbitrary*. While the overall form of a sentence in any given language might seem arbitrary if regarded as a computational system, if looked at as a necessary means of maintaining legibility in the information stream, it looks much less arbitrary. For example, superiority is a WH movement restriction, as in (43), that might seem largely arbitrary.

(43) a. Who bought what?

b. * What did who buy _?

While (43b) is generally considered bad, upon repeated listening, or in a list-pair environment, it becomes more acceptable.

(44) *Scenario: several friends are buying food for a party; the questioner is not interested in exactly who bought what but is more concerned that each person did in fact buy something and what that thing is.* Q: I know that the fenugreek was from Amps and that the tahini butter was from Caroline but I want to know what everyone else brought before I go crazy—tell me, *what* did *who* bring?

This suggests the effects in (43) are perhaps due to clarity of parsing, i.e. a less ideal ordering, and rejected because of this oddity as opposed to an insurmountable problem for parsing. A more complex phrase, as in (45b), shows that superiority violations can be a serious problem for parsing.

(45) a. What microscope did you use _ to view what worms hatch?

b. * What worms did you use what microscope to view _ hatch?

Both approaches attempt to reduce it to a step by step system that assembles from basic components but a system that assumes a discrete computational system—i.e. assumes the form of language primarily takes place inside the head of the speaker—perhaps inevitably begins to regard language as an arbitrary system; because anything is hypothetically possible in the unknown of the brain, it seems that there must be specific limits arbitrarily imposed by the system by something like UG. The view that assumes that language, while incontestably a product of the brain, does not necessarily take its restrictive form from the workings of the brain. That is, if the information stream is at least an equal part of the system, its character reasonably has an influence on the final shape of language.

4.1. Argument Domains

Argument relations (A-relations) are generally considered to be formed by the thematic roles (θ-roles) of verbs acting on noun arguments (REF). As discussed in §2.7 and §3, it does not seem that there exists a generalized and recognizable audible *identifier* necessary for the brain to reliably recognize something like lexical categories; as such, interactions based on the fact that a lexeme belongs to a particular lexical category or encounters the member of another lexical category in and of itself cannot be enough; the same reasonably holds true of thematic roles. That is, if we are to fully dispel any notion of metaphysics in discussion of brain function, we have two choices when it comes to thematic roles: i) verb-noun selection is based on a speaker's prior knowledge of how a given verb interacts with particular nouns, ii) there exists some tangible system of identifiers (or signature) on lexical items.

There is a tendency to denigrate the wealth of data a speaker is exposed to, the thought being that instead of merely having something like a list in one's head, there must be some means of induction; if this data is insufficient then reasonably there must be a reliable means by which a speaker can group or process lexical items according to their thematic roles etc. That is to say, generalizations based on lexical category membership may not be possible without specific knowledge of the given lexeme; i.e. if a speaker has exposure to a lexeme such as 'walk' in isolation there seems to be no reliable way to classify and subsequently use the lexeme; once the lexeme is placed in syntactic context (e.g. 'I *walk*ed'), there is much more information about it. This is not to say that induction is never possible but that, until the lexemes in question are known, generalizing from them is highly problematic; once they are known, it is uncertain how much we can say the brain is proceeding by induction and how much by prior knowledge. Jackendoff (2002) and Mel'cuk (personal communication) both agree that the number of argument relations seem to vary between zero (in cases of semantically null themes such as 'it' for verbs like 'it rains/snows/drizzles') to about four in transaction verbs (e.g. 'agent x rented item y to renter z for period w'); where they disagree

is in terms of the level of generalization available. Jackendoff, as a representative of the Chomskyan view, argues that the generalizations behind theta-role theory, as problematic as they are, are ultimately worth sorting out. Mel'cuk (p.c.; Mel'cuk (1988); Apresjan et al. (1969)) argues that each lexeme has its own character and its own list of arguments. Given my concerns about identifying lexical category membership—and by extension thematic category membership—, it seems that, until a given lexeme is known in some detail (i.e. known in at least one typical context or, exceptionally, identified analytically by a scholar or dictionary), I do not see that it can be processed syntactically. Conversely, once it is known in such detail we cannot say with assurance that we are processing it according to category membership (lexical or thematic)or according to our specific knowledge. In this question, I side with the work of Mel'cuk and suppose that, while there are certainly generalities to be made in the area of thematic (or θ-roles), they are not consistent and therefore not sufficient for processing.

From a lexical point of view, therefore, I argue that there is a wealth of stimulus. This is not inherently in contrast to the Poverty of Stimulus (PoS) hypothesis, which relates to syntactic structures. The PoS was introduced in Chomsky (1965) and suggests two things: i) children have grammatical knowledge far surpassing their exposure to instances of language; ii) therefore there must be a genetically stipulated 'Language Acquisition Device' (LAD) or 'language organ' that, in essence, provides and guides this knowledge on exposure to a sample language. While many critics of this hypothesis focus on the fact that the stimulus 'isn't so impoverished as all that' (see the special double issue of *The Linguistic Review* 2002 volume 19 numbers 1-2 on Poverty of Stimulus for examples), I consider this largely a circular argument and at least for the time being not resolvable either way. If the PoS hypothesis means that children are not exposed to every single conceivable permutation of syntactic structure (lexeme substitution notwithstanding), it is hard to disprove. What I do contest is the necessary condition Chomsky supposed, that the evolution of a LAD was the only answer to this apparent problem.

Instead, I suggest that by the nature of the spoken syntactic stream, there is not enough information in it to allow certain permutations of the grammar (see§2 and §3). In this section I begin to explain that, given these limitations, it is expected that certain forms of the grammar pattern as they do. I begin by discussing why lexical items, when combined, combine in phrasal units (like XPs) and as such can have clauses and sub-clauses.

4.1.1 Child language

Child acquisition data should not be sidelined as a valuable but secondary research programme. In studying a mature grammar, our primary tool is empirical deduction to determine computational ordering. This is often done in isolation from other research on

language and so we continue with certain conceptions of human processing that we inherited from the early days of our field (for example, that derivations are largely serial, see §2.10 and §4.10 for discussion). This is a daunting task, comparable to a theorist in another field studying and making theoretical models of a less than simplex system. If there ever is a point at which a mature system breaks down, this is the point where we must investigate so that we might get a toe-hold into the workings of more complex systems.

For the goal of understanding adult argument relations I argue that it is similarly backwards to determine an adult sentence and try to explain how a child might learn it. Linguistics is unusual in the sciences for starting with the adult (i.e. systemically complex) data and performing a sort of 'reverse engineering' of the theory to account for how children reach adult levels. That is, taking a 'simple sentence' from the adult grammar is still looking at a more complex interaction than we could; modelling a child's grammar after a simple adult phrase instead of the adult phrase after a child's grammar is losing the important differences in the data that are crucial to understanding the adult grammar. Examining the child's defects in this light has created a rich literature but has not to date provided a clear solution. If we take the lead from children and look at how they begin computing argument relations, this will greatly inform our understanding of the adult grammar.

It is well known that when children start combining words they do so with very simple phrases and that these first phrases are two word combinations that are highly similar across languages (Pinker, 1995, 142). What is unknown is what underlying procedure is the cause of this. Contemporary theories argue that they are using what they have developed or what is available of a predetermined syntactic tree structure. The same data can be taken to be signs of syntactic computation guided by their efforts at lexical combination and constrained by their developing control of their articulatory system.

Comparing basic child acquisition data with the discussion in §3.7, I suggest that the fact that children start their computational efforts with the verb is not due to a truncated structure (Rizzi, 2000) or other theories of innate grammar (Borer and Wexler (1987); Pinker (1984)) but rather that the seemingly truncated structure is a result of their early efforts with the composition of lexical categories into XPs. The apparent regularity is a consequence of the apparent priority of certain categories over others. Basic composition is demonstrated by children in various ways, either by linking a verb to a noun or by adjoining modifiers to a noun or a verb. Where children fail is composing more than one or two items at a time. This can be explained if we do not assume an innate structure that must be 'activated' but rather argument relations that develop one step at a time.

Though at their one word stage children may start with nouns, they are not able to compose anything like an adult sentence with only nouns. Once they start combining multiple entities and progress past the declaration of 'cat' or 'man', verbs take on a central impor-

tance in the composition of the phrase. This is not to suggest that there is necessarily a mental predisposition towards verbs but that if we want to organize and express statements about the world the verb becomes very central[1]. For this and possibly other reasons, given that they can only combine a few items together, children—and we may take this as indication that we do the same—consider the verb to be a pivotal element in sentence construction. They seem to take the verb as a starting point but rarely are able to combine both an external and internal argument of a verb when they begin speaking (REF). Early child grammar shows that children seem able to compose a verb with one other argument or a noun with a modifier but not a verb with two arguments or a verb and noun with its modifiers (REF, EXAMPLE). We must remember that computational powers are limited and choices must be made.

However we want to shape this picture, we must determine why it is that, whatever the limitations that children have are, they choose to express the verb and the object. This may well fall out of a situation where the object must be specified when the subject is most likely first or second person[2]. Modern Hebrew, for example, is a null subject language (aka. *pro*-drop) but only allows null subjects in first and second person (Yosef Grodzinsky, p.c.) and in Spanish null subjects in adult language are very free in first and second persons but much less used in third person constructions (Lipski (unpublished))[3] Therefore, given the limited combinatorial abilities of children, i.e. the ability to semantically combine only one argument with a verb, it is not surprising that they choose the object, when the subject is possible to derive from the local discourse environment. As Stephen Crain is often able to demonstrate (Crain et al., 1996), children do not necessarily *lack* any particular aspect of

[1]In fact, when modals and aspect are introduced into the system these seem to take precedence over verbs; we might in fact have a hierarchy of saliency such as: NOUN > VERB > MOOD where children first focus on objects when they start to speak, then progress to events and finally are able to integrate mood and aspect. For those who doubt that these concepts can be understood or expressed without a pre-existing structure, I present this scenario: a small bird or mouse perceives a visual or auditory OBJECT $< hawk >$ or $< cry\ of\ hawk >$ and then sees an EVENT $< dive\ of\ hawk >$ and translates this into a MOOD $< danger >$. That mood is the most abstract does not presuppose a way of coding this as the most abstract in human language. Rather, it is inherently more abstract and the coding in human language simply reflects this.

[2]Which is to say that while the object might be contextually salient (and according to Fisher et al. (1994) this seems to be case for acquiring the lexical items; see discussion in §3.5.1), the object will in most cases vary more than the subject. Assuming, like Hebrew and Spanish, for instance that null subjects are licensed above all for first and second second person (Cf. CHILDES), they will then be more conditioned for contextual reconstruction than the variety of objects that come into consideration.

[3]This is not to say that the occurrence of null subjects in adult syntax is the same as in child syntax. Borer and Rohrbacher (forthcoming) discusses the fact that children acquiring languages disallowing null subjects will omit subjects in non-finite clauses but rarely in finite clauses but they state that "it not clear that in these cases there are specific agreement features associated with the null subjects" (17). See Roeper and Rohrbacher (1994) for more discussion. I argue that adult null subject languages permit the null subject for the very reason that the ϕ-features are apparent on the verbal morphology. For cases such as Chinese where there is no morphology at all (cf.Huang (1984)), one might regard the highly restricted word order or contextual information as a means of construal. Contextual information is apparently the licensing factor for polite forms in languages such as Spanish (etc.) where the second person polite form exactly matches the third person morphology.

adult computational abilities. Nonetheless, when they do not have all the support provided by explicitly clear discourse environments, they are limited in their abilities to put together a number of the grammatical functions they are able to negotiate either in isolation or in more ideal conditions. All of this aside, once we agree that there can be a deficit, we have to ask why it is consistently the verb that is salvaged. That it is not any particular lexical element *per se* that cannot be expressed but a general computational limitation is also suggested by children's ability to combine modifiers in isolation (e.g. 'black dog' or 'fat man') but not add them in a sentence level effort (e.g. 'The black dog bit the fat but tall man').

Often the most contextually salient items (me, you) are dropped in favour of something that must be said if it is to be understood. We need to examine what is salvaged in the case of language breakdown. The parameter hypothesis or the maturational hypothesis within a larger assumption about parameters is descriptive only and does not pretend to explain *why* it is that verbs and their internal arguments are what 'survive' these instances of sub-adult performance.

The current view is that, as A-structure becomes more complex, more 'nodes' are available in the child's syntactic tree, and consequently, the grammar is also informed of—or matures to incorporate—various stipulations (e.g. 'do not extract out of a clause serving as subject to another clause'). I argue that these stipulations and the riddle about the maturing syntactic tree are explained if we regard these as effects of the maturation of the general computational abilities of the child—if we look at how the child goes about computing or 'composing' these argument relations. Assuming the verb as a starting point for argument relations solves many problems in the description of the adult grammar and the interrelations of various clauses. In my description I refrain from terms such as CP or IP for reasons that become clear; instead I describe these computational or compositional domains as *Argument domains* (A-domains).

4.2. Basic Composition

As a means of arguing for an innate linguistic structure, generativists occasionally have pointed out that there are certain constructions that children simply never say. This is taken to be evidence in favour of the innatist hypothesis. For example, Pinker (1994) points out that children never experiment with question formation by speaking backwards[4] (e.g. "Built Jack that house the this is?" (Pinker, 1994, 234)) or by ending with a determiner (EG) (etc.). While it is worth discussing why we never end with a determiner (etc.), if we respect the sense of the lexical categories as well as what meaning the children might be

[4]Note that Pinker's example in Language Instinct of backwards question formation in language is formed badly: It should be: Built Jack that house the is this? Not as he has it in Language Instinct (this observation is pointed out in Sampson (1997) page 122). If this typo shows anything, perhaps it indicates how prone we are to fixed patterns.

trying to express, we do not need to go too far afield to start to see a pattern in their speech. That is to say, if we accept the Full Competence hypothesis (Poeppel and Wexler (1993); Borer and Rohrbacher (forthcoming) etc.), why is it that we never see phrases as in (46) following attested[5]?

(46) a. Daddy <*goes*> to <*work*> today.

b. Tomorrow my <*brother*> <*plays*>soccer?

c. Mommy does the bird <*sing*> in the morning?

d. Why does the <*man*> look like a <*monkey*>?

At once these sentences seem absurd. Nonetheless, they are not completely random is their sense. Some have lost their (non-copular) verbs and some their nouns. This is generally counter to principles of composition in generative theory by means of the *projection principle* (Chomsky, 1981, 34–48), whereby heads project their category; from this, seemingly logically, verbs must head their projections (VPs) and nouns must head theirs (NPs). The trouble is that this is a theory internal result; as they are defined, VPs must have a V^0 as a head. It does not answer the basic question, why should heads be present? In fact it is worth noting that the theory I am advocating is more restrictive than a generalized generative theory; that is to say, while both theories address the data (as they must), the generative theory is based on observation of the existent and expressed categories. I see no fundamental reason why it could not be modified to accommodate different empirical facts. The theory I propose has no clear means of modification to accommodate different facts. We can take it as a given that every theory has to explain the empirical facts. But it is worth asking if a theory is explaining these facts from basic principles or as some kind of counter-instance, anomaly or deviation from its basic assumptions. If the explanation does not come from basic principles, one might ask if it is more descriptive than explanatory. The existent data is consistent with both theories; if functional categories and not lexical heads made up the core of early child language data, this would only be consistent with a generative theory.

Following from §3, I suggest that the order of basic composition may be something not unlike what we see in phrase structure rewrite rules (except that we do not have a concern with functional categories), as in (47).

(47) a. nouns may take (adjective) modifiers for specific detail

b. verbs compose with nouns to attain reference

i. in transitive verbs, priority is given to the object because the subject is often able to be reconstructed from context

[5]Where the greyed out words are not spoken but covertly present.

c. both (47a) and (47b) may be mutually exclusive in early stages of child language, presumably because of performance issues

4.3. The Root Infinitive

As a sample of early efforts at composition, we can take a look at a commonly studied form of composition, the root infinitive. As discussed in Hoekstra and Hyams (1998), it is an established observation that children acquiring German, Dutch, French, Swedish (etc.) go through a stage when they use infinitive verbs in certain contexts, called 'root infinitives' (term originally from Rizzi (1994)), as in (48).

(48) a. Thorstn das haben
Thorstn that have-inf (German, Poeppel & Wexler, 1993)

b. Papa schoenen wassen
Daddy shoes wash-inf (Dutch, Weverink, 1989)

c. Michel dormir
Michel sleep-inf (French, Pierce, 1992)

d. Jag ocksåhoppa där ådär.
I also hop-inf. there and.there (Swedish, Santelmann, 1995)

This stage is also called the Optional Infinitive stage by Wexler (1994) because during the root infinitive stage children also produce correctly inflected finite clauses, as in (49).

(49) a. Ça tourne pas
that turns not

b. Pas tomber bébé
not fall baby

Pierce (1992)

The tokens in (49) occur often during the same elicitation sessions so the root infinitives cannot be considered a distinct developmental period.

In passing it should be noted that there is a suitable explanation to the apparent 'choice' of root infinitive versus finite verbs on my view. While I cannot discuss all the languages in (48), if the French data is representative, there may be a plausible explanation to the phenomenon. Regardless of theoretical assumption, it seems reasonable to say that child language is at base an imitation of adult forms. Examining the data in (49) and comparing it with French infinitival forms involving negation, the data does not seem so strange. That is to say, when a complementary infinitive is negative, the negation (ne pas, ne jamais etc.) is normally placed to the left of the present of the infinitive or in front of the auxiliary for the past infinitive, as in (50).

(50) a. J'ai envie de ne rien faire pendant dix jours

b. Il regrette de ne pas pouvoir chanter

c. Nous avons décidé de ne plus suivre ce régime

d. Elle est triste de ne pas avoir gagné la course

(REF)

With this in mind, the fact that verbs occurring to the right of the negative are non-finite and verbs occurring to the left are finite (pace Pollock (1989)), once again is an imitation of adult form[6]. One might suppose that it is quite possible that the children fail to recognize the context when they should be using one form or the other, rather than hypothesizing that the verb moves to check ϕ-features.

Nonetheless Hoekstra and Hyams (1998) have a proposal regarding root infinitives that may be applicable here. They state that the interpretative properties of root infinitives have not been much investigated. They propose (51).

(51) The Eventivity Constraint (EC)
RIs are restricted to event-denoting predicates

By this they mean that "there seems to be a constraint on the aspectual nature of the verbs occurring in RI-constructions, viz. only eventive verbs are allowed in such constructions, whereas stative predicates occurring during this same period typically require finiteness". Following from this, they observe, "RIs typically do not get a deictic tense interpretation, but rather receive a modal interpretation" which they state as (52).

(52) The modal reference effect (MRE)
With overwhelming frequency, RIs have modal interpretations

They report on the findings of Ferdinand (1996) that there is an 'eventivity' constraint in early French. Their list of stative verbs in French that *do not* occur in root infinitives (but only in finite forms) are as in (53).

(53) Stative verbs in early French: finite only
avoir(have),être (be), s'appeler (be called), manquer (be absent, lack), vouloir (want),

[6]I am speaking about the 'form' of the phrase not specific verb-negation combinations. Clearly children are not simply recording what they hear and repeating it back; there must a principle of substitution at play. This does not necessarily suppose an innatist solution. For example, although the fact that, if we see a face, we can correlate it with other faces, could be answered with or without an innatist theory, we have this ability with objects for which we clearly we have no innate knowledge of. We could include a list of manmade objects such as cars or cell phones, things that have been with us for so short a time in our evolutionary history that there is no plausible means of attributing innate knowledge to them. So if we can recognize and make a correlation between one known instance of a telephone to another, i.e. substitute some visual knowledge of one for knowledge of another, it stands to reason that we can do this for audible knowledge. Tests reported in Pinker (1994) demonstrate this with made up words such as *wug*. It is clear that we do not have to memorize every last instance of every last verb-neg combination. See discussion in §4.9.

> croire (believe), plaire (please), aimer (love), adorer (adore), espérer (hope), savoir (know), se souvenir (remember), devoir (must), falloir (be necessary), pouvoir (can), *aspectual* aller (go)

They point out that a compelling example of the stative/non-stative division is that the verb 'aller' occurs as a root infinitive but only in its main verb sense (an observation they credit to Ferdinand (1996)).

In comparing the child root infinitives with adult infinitive uses in these languages, Hoekstra and Hyams (1998) point out that

> while finite utterances describe actual states of affairs, RIs do not refer to actual eventualities, but to eventualities that are not realized, and are therefore interpreted as statements of desire with respect to these eventualities. Importantly, children's RIs are very similar to RIs in adult language in this respect. Adult RIs have a much more restricted use, but to the extent that they occur, they have a similar [−realized] aspectual value, with an imperative or counterfactual meaning

They quote data from Wijnen (1996) comparing how adults use root infinitives, in (54).

(54) a. jussives

> Hier geen fietsen plaatsen!
> here no bicycles place-inf
>
> 'Don't put bicycles here'

b. Mad Magazine sentence

> Jan met mijn zus trouwen?! Dat nooit
> John (with) my sister marry-inf. That never

They comment that jussives in adult grammar are the most similar to the sort of root infinitives found in child language and like most child root infinitives, they imply deontic modality. The data in (54b) also communicate non-realized eventualities. Specifically, the possibility of the eventuality is raised in the first phrase to be commented upon in the second. This is to suggest that the modal interpretation of child root infinitives is also a feature of adult infinitives. In this sense, the child speaker is not mistaking the context at all but using root infinitives as adults use infinitives. The difference is therefore not that root infinitives (or their analogues) are barred from adult grammar but that they have a much more restricted use (Hoekstra and Hyams, 1998§5); Hoekstra & Hyams claim that the

differences between adult and child use therefore is not a grammatical difference between populations but "a difference at the interface of grammar and discourse" and that "In the adult system the grammar generally wins out, while in the child's system there is a greater reliance on discourse, and presuppositional information".

The above discussion of root infinitives suggests that (pace Pollock (1989)) a movement analysis is not necessary to explain root infinitive data; the form taken is perhaps a 'choice' of two distinct forms and not the permutation through movement of what is basically one form.

4.4. Domain of the Verb: A-domains

If the verb is the of primary importance for the composition of sentences (see §3.7 and Fisher et al. (1994)), there should be further evidence for this in other areas of grammar. I investigate the idea that if a verb is the central element in argument relations then this should help explain the distinct character of clausal islands and adjuncts.

My argument can be summarized as follows. A verb in a main clause in effect does not care about other verbs but only its own arguments. If one of its arguments is a clause in its own right, the verb of the main clause will not interpret this smaller clause according to its inner composition but in terms of the matrix verb's requirement. I.e. a clausal subject will (barring exceptions) be interpreted as an impenetrable whole. Smaller A-domains may generate along similar lines, e.g. adverbs and adjectives as well as infinite preposition clauses.

4.4.1 The Implication for Semantic Composition

If computation of argument relations proceeds as suggested in the previous section and §3.7, this implies some interesting results for semantic composition.

Standard theories of semantic composition proceed with the composition of elements that are locally linear, e.g. noun with its modifiers, whether they proceed 'top-down' or 'bottom-up' ((Heim and Kratzer, 1998, 99ff.) without particular regard for composing with the verb of the sentence. E.g. in the sentence 'The red apples are rotting' the set of all red objects will compose with the set of all apples to find the intersection containing 'red apples'; this set will compose with the set of all objects that are rotting to find the intersection which would be the set of red apples that rot.

It is curious that a system of semantics cannot distinguish between some very basic properties of the world, that there are *event* entities, i.e. objects that are representations of—or exist in—time-space; and *non-event* entities, i.e. objects that exist in physical space and can be tracked by their relation to time space and to each other or are abstractions of objects that can be so described.

A theory where the verb first combines with its arguments and is then conditioned by the modifiers of the argument (or verb) provides interesting solutions for certain problems in the semantic literature. Currently a sentence such as 'The former president is a liar' has the problematic composition of the set of all objects former with the set of all presidents. This results in an empty set before the verb has been able to compose with the subject argument. A verb first analysis avoids this problem but runs into an apparent complication of a modifier applying to entities it is not meant to modify, e.g. 'red' possibly applying to 'rotting' at the same time as applying to 'apples'.

4.4.2 Are all Babies Russians?

The gains described in (4.4.1) are valuable but as noted not without problems to resolve. Among the problems is the semantics of seemingly semantically null entities such as the copula. Russian is an example of a language that seems to be a counter-example to a verb-first composition scheme. To this end I will investigate the function of the copula in languages like English where it carries no significant meaning. Even in English the copula seems not always to be present in the cases of small clauses and preposition phrases; an example is a sentence like 'John ate the sandwich in the kitchen' where the meaning is something like one of the following.

(55) John ate the sandwich in the kitchen

a. John ate the sandwich & John (was) in the kitchen.

b. John ate the sandwich & sandwich (was) in the kitchen.

The question seems to be this: do we proceed from most concrete (nouns) in our one word utterances, proceed to more abstract events (verbs) once we start to compose sentences and then once we start to use aspectual and modal systems consider 'mood' most salient when it is present[7].

***

Perhaps the apparent elision of the copula is related to the same fact that prevents statives in root infinitives in child grammar [SEE HOEK AND HYAMS] and that allows them to not be overtly expressed in languages such as Russian.

4.5. Acquiring the non-Arbitrary

If the computation of argument relations between words falls out in a plausibly derivable way, this has serious implications for the *Poverty of Stimulus* (PoS) hypothesis. Chomsky

[7]Is it possible to investigate this with experimental methods testing what is most salient in given circumstances? The aspectual system seems to be the most complex in many languages, something that many L2 speakers never fully master; perhaps this points to the centrality of aspect.

(1965) proposed this to account for the fact that children never make certain errors. A central argument of the PoS hypothesis starts with the data in (56).

(56) a. Mary is home.

b. Is Mary <*is*> home?

Advocates of the PoS hypothesis have suggested that the rules considered by a biologically uninformed child might be as follows[8].

(57) a. Move the main clause Aux to the front

b. Move the leftmost Aux to the front

c. Move any Aux to the front

In the case of the data in (56), any of these rules would work. PoS suggests that the child will have problems when the new rules are applied to the data in (58) below.

(58) The man who is tall will leave now

Under the PoS hypothesis, the rules in (57) will result in the child variably producing either of the following:

(59) a. Will the man who is tall <*will*> leave now?

b. * Is the man who <*is*> tall will leave now?

That the second of these is not attested has been presented as evidence that the child has a biological predisposition to knowing a matrix auxiliary from an embedded clause auxiliary (Chomsky (1965); Boeckx and Hornstein (2003); Lasnik and Uriagereka (2002))[9].

This assumes a strictly serial view of clauses, that all auxiliaries contained by the greater clause are open for consideration by relations within it. From a top down or final product point of view we may see no inherent distinction between auxiliaries that pertain to the main clause and those that are within adjunct clauses. Nonetheless if argument relations are built from requirements of composition, whether of the main clause or adjunct clauses, these normally exclude elements in other clauses according to the simplest principle of locality of search. Judging from child speech errors (namely speaking in verb-centric fragments

[8]While they talk of 'rules', I suggest it is better to discuss this in terms of patterns; a pattern that is only half used still fulfils its purpose, a rule only half used is somehow contravened.

[9]Another common though less convincing argument involves data such as: i) John believes that Bill kissed Mary ⇒ ii) Who does John believe that <*who*> kissed Mary? (Atkinson, 1992, 41). This data isn't as terrible as sometimes described and is in any case explained by the demands of punctuation; in contemporary English we often drop 'that' complementizers when they are not needed; it may seem odd to some speakers to have one here but mere oddity can be attributed to lack of familiarity and doesn't necessarily have anything to do with rules of grammar.

of clauses), we might consider the idea that children take the piecemeal composition of clauses very seriously. This is not to suggest that we *always* compose everything we say but that *when we compose our syntax, we do so from the point of view of clause building not sentence building*. Importantly, this includes the content of clausal adjuncts. The PoS hypothesis dates from Chomsky (1965) and was proposed without our current understanding of adjuncts and relations between lexical items. If we can explain this data using lexical composition as the principle organizer of argument relations we can look at reducing our assumed biological precondition for language acquisition and therefore our model of how the adult grammar works.

Considering the 'rules' in (57), it is interesting to note that what Boeckx and Hornstein (2003) do not suggest is that the movement is constrained by some form of economy and locality and this is provided by the shortest move. Here of course shortest move cannot mean (57b) but it could very reasonably mean, *thinking only about the clause in question* (and therefore ignoring adjunction and relative clauses), move the most local Aux to the front. I.e. move the Aux that determines the clause to the front. This provides a non-stipulative and natural explanation of the relation of adjuncts to main clauses. In this way, it is reasonable that the child does not need to 'know' the difference between main and adjunct clauses but simply that if a clause in question is being considered there is no reason the child would look outside of the clause. In this way, the child does not need to keep track of adjunction differently than main clauses except that they don't seem to be as important for most operations most of the time; how they are composed is much the same. If composed along the terms that I am suggesting then [the man] will be related to [will leave] before anything else and the child will 'search' for unity is such terms. [who] introduces a break of some important sort (to the adult and to the 'learning' child) and so this constituent must be ignored for the time being. This fits in with other Generativist principles and therefore a Generativist principle need not necessarily be one that insists biology is involved[10].

The principle of 'most local search' is central to generative arguments under the Minimalist Program and is provided in the same text as arguments for the necessity of biological rules:

> The motivations of vertical minimalism are visible in proposals that argue that the grammars do the least work necessary to produce objects usable by the sound/meaning interfaces. So, for example, if something must move to meet some requirement then the movement must be the shortest possible, or if some

[10]While it is a hypothesis (and a sometimes tempting one) it should not so readily be taken as the null hypothesis and is not needed for this question. Indeed, instead of being so quick to shrug off the argument against a biologically predetermined aptitude for language acquisition, the Generativists might play devil's advocate and see if they cannot work up an argument against their own working hypothesis. This is what I attempt to do in this chapter.

> requirement must be satisfied it must be so satisfied by the first available expression that can do so (see Collins 1997, Kitahara 1997). If a search is required then the system is designed to insure that the search is optimal, that the relevant information is easy to get to and the relevant operations easy to implement. (Boeckx and Hornstein (2003), 21)

All that remains is to agree that *clausal adjuncts are clauses* (subordinate or otherwise). As such they would be excluded from the most local search by a child first learning the language. *Presumably the domain determined by what is needed by the verb limits the seach.* To find evidence that adjuncts are somehow distinct is not hard and the idea that they would necessarily be excluded from the most local search follows from the principles of argument relations I am advocating[11].

In standard Generative terminology adjuncts are considered A-bar positions and as such they are outside 'main clause' argument relations. They could consequently be defined in the sense that argument relations do not 'care' about their adjuncts but that the adjuncts necessarily interleave with or 'attach' to arguments because they modify them. The size of the adjunct does not seem to affect this and adjectives, adverbs and various types of modifying (subordinate, relative) clauses are all quite unimportant for for argument relations. This is nothing new from the point of view of argument relations but it is important if we are trying to understand this in terms of a discrete biologically endowed structure. Of course adjuncts might be described as being on another dimension of sorts ((Uriagereka, 2002, 278–280) and references therein) but this does not answer the question of how to integrate them into a standard Generative conception of grammar. This problem does not seem so daunting if we look at the interrelations of the argument relations. Nonetheless if these empirical arguments can be accepted in isolation from that underlying assumption, we can now turn to the basis of the acquisition argument and see where that leads.

4.6. Robustness of the mechanism

As pointed out in Boeckx and Hornstein (2003), "Any learning system will have to be supple enough to ignore noise in the data" (12). This seems a excellent postulate and one I would like to explore as regards my proposal to date.

Chomsky using metaphors of the 50s and 60s suggested that some kind of computation was at play when we worked out the syntax of our sentences. This metaphor has been kept at the core of the theoretical model and continues to be central to the theory; recently, e.g. in Chomsky (2000b) and Chomsky (2001), Chomsky has added terminology such as

[11] For example, Abels (2003) points out that Chomsky (1995b) follows Kayne (1994) in stipulating that "the sisters of specifiers/adjuncts do not participate in c-command relations" (86). Abels furthermore mentions that his own proposal tries to give "a principled reason why this is so" for "moved specifiers/adjuncts" but that in "Kayne's and Chomsky's systems this property is stipulated" (86).

'crashed' derivations apparently taken directly from current computer jargon. The idea behind *phases*, while the term perhaps is simply meant to indicate successive cycles, has also been discussed from a computational point of view in Chomsky (2004) among other works: "What objects constitute phases? They should be as small as possible, to minimize computational load" (17).

Nonetheless our only real 'computers'—the ones we know today, are crude and unreliable machines that, furthermore, need to be given instructions in order to work. If this sense of computation is what we mean when we talk about syntactic computations then this is a serious problem for UG. I know of no other detailed sense of the word, i.e. if we are not talking about the details of the machines we have made then the only sense we have of what it is to 'compute' is what one does in one's head when, for example, someone works out the algorithm of long division; because we have no exact description or representation of what this is, we can only say with precision that either 'to compute' means to act as our manmade computers or else to act in a way we do not currently understand. Why is our computational system not more fragile than it seems? While people make performance mistakes, the errors do not seem to be of the level expected by synthetic computers (that when they make errors sometimes produce gibberish (etc.)); Pfau (2000) lists the following general areas of speech errors: Anti-Agreement (i.e. "various kinds of slips of the tongue—exchanges, perseverations, anticipations, and blends"); Feature Mismatch between Subject and Verb; Proximity Concord; Feature Mismatch within DP; Errors of Subcategorization; Accommodation and Stranding; Affix Errors. What are not reported are cases where we produce anything like the errors made by our synthetic computers. Why do we have something so robust that even its mistakes do not upset the system? If nouns and verbs and quantifiers are some built-in property of the workings of the mind, then the brain might just 'capture' them when they are identified. But then (as discussed in §2.7) why are there languages without these elements and what is it that would trigger the 'capture'? Strange that it would be something as slippery as meaning; more likely it might be some tangibly audible spike or a certain waveform than 'something that scopes over its variables'. This is not to say that the brain operates in anything like how our computers operate—in fact I suggest this is almost definitely not the case (see Hawkins (2004) for interesting discussion) but it must be understood that to date we have no detailed model (working or theoretical) of a computational system that is robust. As Artificial Intelligence theoretician and software programmer Jaron Lanier points out, "Software is brittle ... If every little thing isn't perfect, it breaks. We have to have an honest appreciation for how little progress we've made in this area" (Lanier (2000)).

Clearly this is the wrong model for the robust system of human language and mind. While a better model is beyond the scope of this essay, in the next section I consider a different computational protocol, one that has immediate implications for syntactic theory.

4.7. Poverty of Stimulus & Negative data

Legate and Yang (2002) (see also Boeckx and Hornstein (2003), Yang (2002)) discuss arguments from Sampson (1989) and Pullum and Scholz (2002) regarding the quantifiable nature of linguistic data. In fact, it is quite possible to quantify the rareness of the data where a verbal auxiliary appears in an adjunct linearly prior to a matrix clause adjunct. Sampson (1989) and Pullum and Scholz (2002) use Wall Street Journal corpus as a source of apparently typical English; of the initial five hundred sentences analysed, five (i.e. one percent) are of the form as in (58), repeated in (60).

(60) The man who is tall will leave now

Nonetheless, to argue that this is a relatively great number of tokens or evidence of impoverishment is an essentially circular argument.

Instead, if the child understands that the argument relations hold of a given simple sentence and later learns that modifiers can be added to the various elements of this simple clause, they need only proceed in a stepwise fashion to 'learn' to incorporate adjunct clauses. As illustrated in (61), it is fully possible to modify most of the lexemes in the sentence without disturbing the argument relations.

(61) a. The man will leave

b. The tall man will leave

c. The man will leave quickly

d. The man will not leave

To determine what physiological function pertains to the general mental abilities of selection, choice or identification is beyond the scope of this essay and linguistic theory in general. Maintaining reference to a chosen sentence (or object) is not the question we are attempting to resolve. Once the child has the volition to speak a certain sentence, it becomes a question of how the child might modify it—but there is no reason to presume that the child would lose reference to the original sentence. That is to say, if a child is able to generate a simple sentence, α, as in (61a)—and through some undetermined physiological function maintain this sentence α in mind—it is not necessarily a great mystery that the child does not start mixing auxiliary verbs from sentence α with those of another clause β generated to annotate sentence α. Only assuming a fully serial model will this be a question. If a given subordinate clause β is generated in a non-serial relation to α and obeys its own argument relations, to interleave it prior to speaking it is less of a mystery than supposed. This is to say, if the child can recognize the principle object of desire (the matrix clause) and distinguish it from modifiers to this (the adjunct clause), then it follows that the child

is well equipped to handle this more complex data before noting the 1% usage of the Wall Street Journal.

4.8. The importance of the Poverty of Stimulus

It is difficult to overestimate the importance of the Poverty of Stimulus hypothesis (PoS) in the study of language. Prior to Chomsky's hypothesis (in Chomsky (1965)), hypotheses such as 'motherese' and behaviorism (Skinner (1957)) and 'culture' (Sampson (1997)) greatly underestimated the problem facing the child. Nonetheless, this problem does not necessitate a hypothesis of biological predetermination of the kind that often is supposed. Just as a child learns to walk taking one step at a time, so too can the child learn to talk by connecting one lexical item at a time and, by developing the necessary mental co-ordination to 'juggle' a number of these, eventually learn to talk at adult levels.

That it takes children more time to start using more complex sentences can be explained under either the theory under development here, i.e. because the child has to understand the argument relations better and be able to cope with interleaving adjunct clauses, or according to theories such as the root-node hypothesis of Rizzi (1994). The root node hypothesis, put simply, supposes that the available developing tree structure can effectively telescope and allow the child access to various levels of inflectional nodes. Rizzi's theory is unfortunately powerful and quite possibly equally unprovable as un-disprovable. That is to say, while it is descriptive, it is empirically unsatisfying in a number of ways. One could allow that the child's use of such a developing tree structure (if otherwise demonstrated to exist) would be tentative and subject to experimental probing as though trying out a pair of wings. That the CHILDES data is so variable from one token to another is nonetheless curious. (EXAMPLE) If comparison to other developing abilities of the child, one might reasonably expect some warm-up time and then some plateau of competence during a session. If we compare it to a human child's ability at learning to walk we do see something of this pattern, where there is a period of 'getting up to speed' with yesterday's progress and then some plateau and maybe a falling off once fatigue sets in. Might there not be a similar pattern in speech if it is a matter of learning to 'use' higher syntactic nodes? On my hypothesis, the child must parse each sentence on its own and deal with possibly confusing factors such as how to map the argument relations given lexical items of variable familiarity[12].

Both hypotheses need to be able to account for the attested errors. On the view I am advocating this falls out as a clear and inevitable product of learning. On the UG view the theory has to suddenly undermine its principles of inevitable correctness and suggest

[12]This last point might be equally valid for a 'biological' development theory but, because I am assuming that a discrete system necessarily needs access to word category information, the category of the item might help more with processing than if the words are just sounds and meanings to relate to one another.

why the child data fluctuates so rapidly. A null hypothesis needs not only to be a possible solution but always the least improbable one. If both a biologically endowed UG and a theory of lexical interactions satisfy the demands of the data, the second requires less fanciful predictions and should be considered not only as an important re-evaluation of the now long-standing biological view, but also quite possibly the *new* null hypothesis.

4.9. Generalizing from instances

Without the guiding hand of Chomsky's Universal Grammar or an *un*-impoverished source of data the child can still reasonably understand and produce sentences such as 'The man who is tall will leave'. Following from the analysis in Chapter 3, language can be understood to be compositional without biological stipulation that it should be so. On this view it is sufficient to have encountered and become habituated to declarative sentences of (mostly) known words to be able to identify another such declarative sentence of mostly known words. If this is true, it can be said that they have begun to understand argument relations and how sentences are formed. At this point, once they have seen and become habituated to interleaving adjuncts into declarative sentences, they can be said to understand how adjuncts relate within these argument relations. We can then say that the child could produce a token sentence without having memorized it exactly. At this point, a variety of sentences following generally similar patterns might be attempted. It follows then that, if argument relations within clausal adjuncts compose according to the same principles of interaction as in main clause argument relations, the clausal adjunct will not be produced in series with the main clause argument relations. From this it is sufficient that question formation by auxiliary raising choose an auxiliary from the targeted clause and not adjacent clauses (that, quite possibly, are not interleaved at this point).

To summarize, if from the logic of lexical categories (chapter 3), children understand simple argument relations and from the impoverishment of the syntax (§2.7 and §2.11) we can expect locality of argument relations and movement; if these hold, then we can expect auxiliary raising of the only auxiliary verb in the argument domain of the clause considered[13]. Therefore, the child knows to raise the most local auxiliary verb but also knows to exempt adjuncts from an understanding of argument relations simply because, if argument relations create the clause, they cannot consider auxiliaries that are not within the clause thus created (and an auxiliary verb in another clause cannot be the most local).

There is, in principle, no need for a child to have encountered more than one specimen of a simple sentence, a simple question and of a sentence with clausal adjuncts to create an amalgam of these. It is even possible that the child need *never* have seen them all interact

[13]Or some analogue to this; this line of reasoning tacitly supports the theory of transformations, something that I will overlook for present purposes.

in the same sentence to come to the conclusions that they would all work together. (This happens in other contexts where it is the rules and not necessarily a particular assembly of rules that is important—the child need only know that rules can be concatenated.)

Essentially we need to review whether we believe that *systems* can have different layers in their organization (e.g. non-serial clauses) and if these layers may actually dictate how a 'search' or generalized relation forming operation may function. That is, can a search function be sensitive to layers and could this cause a system to operate in non-linear ways? If this is possible, does this search have to be dictated by some kind of biological hardwiring? Remember that the biological thesis was only advanced as the *best possible* hypothesis in face of the daunting question of 'what else could it be?' If we have another answer, we should investigate this as fully as possible because it may well lead us to answers in other domains.

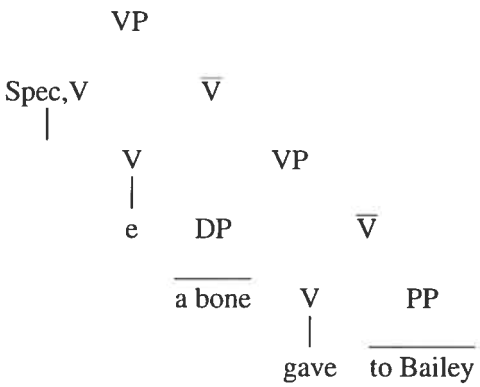
4.10. Serial vs. non-Serial Parsing

From the beginnings of generative grammar, i.e. in Chomsky (1955), the theory has considered it a basic observation that human language is a non-linear phenomenon (see Uriagereka (2002) and references within for some current discussion). Nonetheless, from these same beginnings, human syntax is seen in almost strictly serial terms. By this I mean that, while words or phrases in a derivation might have a canonical or 'underlying' relation with other words or phrases and that these can be permuted or rearranged in certain ways while retaining the same or similar meanings and are as such 'non-linear', they are considered to be part of a schemata such as a syntactic tree that dictates how these items organize, reorganize and otherwise relate to each other as part of a defined *series*[14]. For example, in the syntactic trees in (62), while there may be a transformational relationship between one tree and the other, *each tree in and of itself is a representation of a* series *of lexemes in a presumed syntactic derivation*, i.e. there is little to no presumption that various XPs are computed outside of the derivation and 'flown in' to be added or interleaved with the final spoken token.

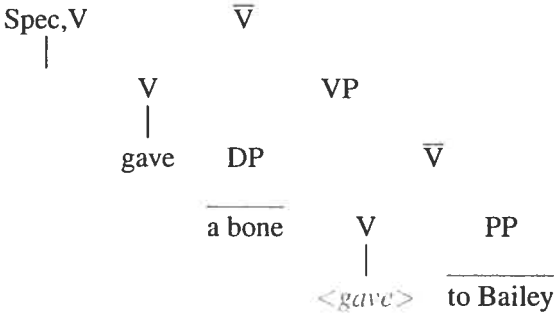
After Larson (1988), with (62) below, where it is assumed that we have the underlying form (62a) and (63a) as well as two possible derivations (62b & 63b)

[14]The term 'serial' is commonly in use in electronics and computing where it refers to a transfer of data by a wire or port as a single stream or sequence of bits or information; alternately, when referring to a processor, it refers to one that runs only a single task at a time (contrasted with, for example, 'multitasking').

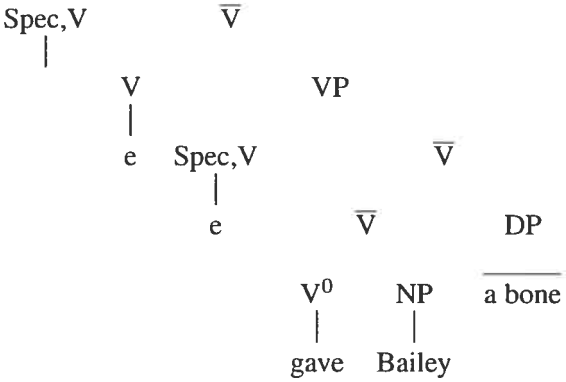
(62) a.


VP
Spec,V
V̄
V
VP
e
DP
V̄
a bone
V
PP
gave
to Bailey

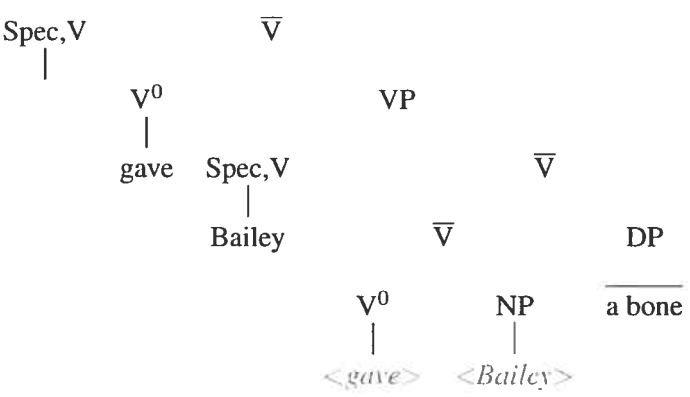
b.


VP
Spec,V
V̄
V
VP
gave
DP
V̄
a bone
V
PP
<gave>
to Bailey

(63) a.


VP
Spec,V
V̄
V
VP
e
Spec,V
V̄
e
V̄
DP
V0
NP
a bone
gave
Bailey

b.


VP
Spec,V
V̄
V0
VP
gave
Spec,V
V̄
Bailey
V̄
DP
V0
NP
a bone
<gave>
<Bailey>

In fact it is just this quality of being 'flown in' that I mean when I refer to a non-serial view of language composition. Commonly in more advanced computing it is necessary that there is some simultaneous performance of various operations and this is often referred to as 'parallel processing' (although the original idea recently comes from that of electrical circuits or components that are connected only to common points at each end and are not connected to each other in sequence). To date, I know of little discussion of this in terms of processing of syntactic derivations. What work I do know of is in the realm of experimental psychology of language, for example in work by Neal Pearlmutter and Edward Gibson. Nonetheless, their work seems to concern not the piecemeal assembly of clauses in parallel or at least non-serial fashion but the possibility that the brain maintains more than one structural representation of certain ambiguous sentences. To distinguish my conception of a the human parser from their usage, I do not use the term parallel but rather *non-serial*; this term also has the advantage of not making judgements on whether or not these sentence strings are in fact processed in parallel or perhaps assembled in other non-serial ways; we await empirical findings before this distinction can be further refined.

One thing to recognize is that a system that demands that everything is computed serially must have such things as 'main clause' or 'adjunct' specified, whereas a system that computes these independently and in non-serial fashion does not have this same inherent struggle to keep these clauses separate. Not keeping them separate clearly leads to garden-path errors. That is, as pointed out in Gibson and Pearlmutter (2000), "It has long been known that the human parser does not retain all possible structural interpretations for an ambiguous input string in parallel, because of the existence of garden-path effects" (231). That is, the potential 'brittleness' mentioned by Lanier (2000), is a very real problem.

4.11. Lebeaux's 'Late Merge' as an example of non-serial syntax

Following from Lebeaux (1988), Chomsky (1993) and Fox (2002) discuss the idea of the 'late merge' of relative clauses and adjuncts in general. The syntactic evidence for this relates to the observations that both *wh*-movement and theories of 'quantifier raising' are are generally not capable of fixing a Condition C binding violation (problematic because of the copy theory of movement, both of which are in general principle adopted here). As long as the Condition C violation is caused by *wh*-movement of an adjunct clause containing a name (Freidin (1986) cited in Fox (2002)).

The data is as in (64).

(64) a. ??/* Someone introduced him$_i$ to every friend of John's$_i$ → [every friend of John's$_i$] someone introduced him$_i$ to t

b. ??/*Guess [which friend of John's$_i$] he$_i$ visited t

c. Guess [which man that John$_i$ likes] he$_i$ visited t

Chomsky (1993) argues that this data is consistent with the copy theory of movement once Lebeaux (1988)'s proposals regarding 'late merge' of relative clauses and adjuncts is accepted. In Chomsky and Fox, these are proposed to be added to a syntactic structure 'countercyclically' (Fox, 2002, 69). The reason that (64c) but not (64a) and (64b) does not violate Condition C is because, according to Chomsky, adding complements countercyclically would violate the Projection Principle.

4.12. Antecedent Contained Ellipsis

The case of Antecedent Contained Ellipsis (ACE, formally known as Antecedent Contained Deletion, ACD) is an interesting one that suggests that an interpretation of the data as essentially serial computation cannot be correct. Recoverability of deletion (Sag, 1976) states that a VP can be unspoken if its meaning is 'recoverable' from the contents of another (previous) clause as in (65)[15][16]

(65) a. I'll vacuum the house today if you will tomorrow

i. [=]I'll vacuum the house today, if you will *vacuum the house* tomorrow

ii. [≠]I'll vacuum the house today, if you will *paint the house* tomorrow

The problem presented by ACE examples is that in the case of a sentence like (66).

(66) $[_{S_1}[_{NP_1}$ She] $[_{VP_1}$ wanted $[_{NP}$ no book $[_{S_2}$ that $[_{NP}$ he] $[_{VP_2}$ did]]]]]

Here the antecedent VP_1 ('wanted no book...') contains the elided VP_2. This is an apparent problem of infinite regress because if the elided VP tries to reconstruct itself from its antecedent it would have a copy of itself in each attempt. Aside from the logical difficulties this presents, the simple demands of reconstruction through identically would fail (Fox, 2002).

Fox (2002) offers a solution to the problem by suggesting that the antecedent is external to the VP. I support this conclusion but argue that the proposed structure is unnecessary and unfounded. The antecedent, I argue, *is* outside of the VP just because the embedded clause is not 'embedded' until spoken output. That is, in (66) the pre-spoken output matrix clause would look more like (67).

[15]Nonetheless, in a discourse heavily focused on who will paint the house tomorrow, (65a-ii) would be perhaps more licit. This does require heavy biasing of the discourse environment though.

[16]For other accounts arguing against rightward elements being 'contained' within leftward elements, see Büring and Hartmann (1997). This paper presents an in depth discussion of embedded clause extraposition (and interactions) showing that in German (and by implication, perhaps other Germanic languages, extraposed clauses are not "hierarchically lower than the material preceding them" which they take to present strong evidence against Kayne (1994) and Haider (1993).

(67) $[_{S_1}[_{NP_1}$ She$]$ $[_{VP_1}$wanted $[_{NP}\ \beta$ book $[_{S_2}\ \alpha]]]$

Here the phrase 'She wanted book' is self contained in terms of its own argument relations but at some point before it is spoken will compose with two other clauses α, β[17] and the order of composition will vary according to the intended reading. If syntactic computation is equivalent to semantic composition then the order of composition would be either: [[no [book]] john wanted] OR [no [[book] john wanted]]. Which it is will be reflected in the reading, i.e. there is a set of books he wanted and a set that she wanted and her desires did not intersect with his; or she had a desire to read different books than he read and so her desires were tempered by his.

Assuming some kind of placeholder or referential unit to be the case for the present discussion, it is important to understand that in terms of the computation the α-, the β- and the 'matrix' clauses will compose without one being contained within the other. If the matrix clause were read 'first', upon reaching α, the instruction is given to the the output parser to refer to clause$_\alpha$ and modify the 'book' accordingly, as with any modifier. The α-clause is never contained but the information in it is free to co-refer with information in another clause as two conjoined clauses might (as in (65)). In another scenario, the clauses may be computed internally and independently but with their internal arguments able to co-refer to each other. Only at spoken output one is 'embedded' in the other.

In the next chapter, we look at impediments to syntactic form due to legibility constraints.

[17]Whether there are, at some level, indicators such as α & β may be a question for inquiry; nonetheless for deletion under identity there cannot be placeholders that would interfere with this. Which is to say, that while there may be placeholders they would have to be indicated at a later point in the computation. That there are more than two tiers to the system is likely necessary and quite expected. see Pfau (2000) for the correlation of DM and speech errors that suggests that categories do play a role in lexical access; this consequently suggests that the shaping of composition is multi-tiered and so likely has a point where the above placeholders would have an effect that does not undermine the needs of 'ellipsis under identity'.

Chapter 5

Limits on the System

5.1. Parsing without Indices

In this chapter I outline the conditions that restrict the grammar because violating them would cause interpretation to fail. As pointed out recently in an editorial in *The Lancet*, until the nineteenth century it was thought that fever was one of the the deadliest diseases that afflicted our bodies; we now recognize that fever is in fact brought on by a number of causes and, although it once seemed to be a distinct entity, is a side-effect of our bodies' response to illness or injury (Lancet, 2006, 705). While syntax, and restrictions on syntax, are currently considered to be due to a monolithic (though not necessarily contiguous) mental module (Chomsky (2005)), might we recognize someday that what we now call Universal Grammar (UG) is due to a number of causes, neural and environmental (etc.)? Perhaps the syntactic forms manifest by UG are—like fever is now known to be—also side-effects of how our brains process information in response to factors present in our articulatory and perceptual systems. Language, by the nature of what it is, must respond to certain restrictions that may well have nothing to do with how our brains operate, or may be partly due to our mental physiology and partly to the nature of information processing.

Instead of presuming to design what must be present in the human brain for language to operate within it, I examine what elements seem to be present in manifest syntax and how these might fit in with a system that relies, at least in part, around discourse between speaker and listener. Although many of the design elements proposed by theorists assume that language operates in a discrete system without reference to the listener, I focus on how listeners could possibly understand the only the level of output available to them, spoken syntax, and discuss how it must be the locus of interpretation. Recently, under the Minimalist Program, the importance of manifest or spoken syntax has fallen out of vogue for conceptual reasons (Chomsky (1995b)), yet if we are to look at the problem from a scientific point of view, it is a grave error to disregard the environment in which our subject of study operates in. In basic science, determining the line between the 'system under examination'

and 'everything else' is in some cases clear and in others never better than a workable approximation. Engineers (and many physicists) can determine their system and perhaps use Newtonian physics with no reference to quantum theory or Einstein's contributions; surveyors can routinely assume the world is flat (Kuhn (1962)). Nonetheless, the working models adopted in this way are conveniences necessary for the practical workings of the subjects under evaluation (e.g. surveying building sites, designing a structurally sound bridge etc.). Constraining one's view in this way is a delicate matter. Sometimes these decisions can be made once and for all time (a surveyor need not consider computing the effects of Relativity on his calculations) and sometimes it has to be decided for each experiment (a chemist has to determine if the reactants in any given experiment are reactive with air). Nonetheless, the purview of *what the system is* is a decision that has to be made with careful consideration at some point in any given field. In the case of linguistics, while our biology clearly allows us to learn language in a way that the biology of a rabbit does not allow it to learn language (see §1.1), to assume from this that all restrictions on language are therefore due to this biological difference is overly hasty reasoning. To consider the problem without reference to listener comprehension is a gross scientific error. With analogy to chemical reactions, I hold that it is highly improbable that the listeners is not 'reactive' with the speaker; normal speakers who proceed without consideration for the comprehension of the listener are considered, at best, confusing, at worst, unintelligible. It is a dubious assertion that language exists without consideration to how it will be interpreted. Why then exclude such a feedback mechanism from a scientific theory describing language?

As discussed briefly in §2.18, even if syntax itself has no learning function or even perception of the effects of its forms on a listener (and it is reasonable not to personify systems), if human speakers were unable to be understood by listeners, *the desire to speak comprehensibly would constrain the system.* This is to say, if there were no mental 'component' barring *wh*-island violations but because of lack of, for example, indices in spoken syntax, listeners failed to follow the resulting syntactic forms, it is quite reasonable *without supposing system intelligence* that speakers would tailor their sentences for listener comprehension[1]. Given that, for the time being and perhaps for all time, we cannot know the inner workings of proposed systems of grammar as in, for example, Chomsky (2000b) or Chomsky (2001) we might consider examining more quotidian causes that might limit syntactic form. Although it is conceivable that a system could operate in isolation and without regard for a listener (where the listener is more of an eavesdropper than intended audience), if we can *use* the principle of listener interpretability to constrain the system then the stipulations to constrain the grammar are that much less arbitrary. Here I assume the impoverishment

[1] Although a dyed in the wool Generativist could easily consider this condition as satisfying an 'interface condition' as proposed in Chomsky (1995b).

of the spoken information stream, as discussed in chapter 2.

While the LF hypothesis suggests that the locus of interpretation is at LF, it becomes a serious question how listeners—presumably without access to this level—can understand spoken syntax[2]. If one presumes that they somehow reconstruct the sentence into their own LF, it is still not clear (if we presume that some kind of 'features' are necessary for movement) how this happens or if it does, what the difference is between 'understanding' a syntactic string enough to reconstruct it and simply interpreting it in its attested position. Furthermore, if the listener is able to either interpret or reconstruct spoken syntax, it is necessary to understand to what degree speakers can or cannot interpret their spoken syntax in its output form. This question raises the possibility that current theories of how indices are interpreted (Fiengo and May (1994) among others) need to be reconsidered. Theories that restrict interpretation to an essentially serial ordering and assume a transformational derivation to order the lexical items in a suitable linear order do not give a clear answer as to why listeners are able to interpret phrases that seem to violate binding principles[3] .

This essay argues against the idea that language is generated by the permutation of a single series of lexemes, that interpretation is only possible if these lexemes attain a fixed relation (i.e. LF) and that spoken syntax is in any way secondary for the purpose of interpretation. While the study of surface forms offers no grand unifying cross-linguistic theory of a computational or transformational grammar[4], it cannot be dismissed as an insignificant level of representation. Although such questions cannot be resolved empirically with our current level of understanding of brain processes, at least for the purposes of *listener* interpretation, there must be means to interpret surface forms. That listeners only have access to surface forms suggests that, as much as language can be understood to have some relation with the process of communication[5], it is likely that there are mental processes that allow

[2]Without means for a listener to reanalyse a given sentence for scope inversion then there would be errors of interpretation all the time. Which is to say that there must be a means to reanalyse and if so why would the speaker not use this means and not go to the business of covert movement when speaking and reanalysis when listening? Reanalysis might be what we mean when we say LF but then this is not necessarily linear (as LF seems to be) and would be more a process of revisiting the overt syntax than letting it continue. I.e. for even the most general purposes of efficiency (if that is a valid concern for language) it seems undesirable and unrealistic to have two mechanisms.

[3]Assuming that listeners are able to incorporate a given phrase into their own derivation for interpretation creates more questions than answers: if movement to further positions in a derivation requires something along the line of Case or feature checking it is unclear how a listener's transformational derivation would involve such transformations *without benefit of the Case or feature function*. If these features ever exist, they are, from the perspective of the listener, inaudible; whether or not this is because they are deleted according to the theories that postulate needing them (e.g. Chomsky (1995b), Chomsky (2000b) Chomsky (2001)). In any case, if the listener does not need these features to facilitate the movement to LF, argued to be necessary for interpretation, it seems odd that the speaker would need them.

[4]See §3.6.

[5]This assumption apparently goes against recent suggestions by steadfast 'nativists' such as Noam Chomsky who has suggested that his hypothetical language organ might have first evolved for the purpose of *beauty*, analogous to a bird's plumage rather than for lay assumptions about communication. The implications of a language organ hypothesis certainly makes this a possibility. I will proceed without such assumptions and

for the interpretation of a syntactic string from only the surface form.

This essay holds that a number of proposals regarding the transformational possibility of language are unproved and very possibly unnecessary. Furthermore, I suggest that levels of representation such as Deep Structure (DS), Logical Form (LF), Phonetic Form (PF) should only be entertained if it can be shown that a theory that restricts itself to spoken syntax (analogous to Surface Structure (SS)) cannot possibly be right. I hold spoken syntax as prior to these other proposals on the grounds that this is our only attested level of representation. A theory that attempts to dismiss the importance of spoken forms out of a "logical conceptual necessity" must account for the reality of listener interpretation; that is, desires for cross-linguistic regularity (one of the bases for the principle of LF, see §2.4) must be considered secondary to principles of listener interpretation.

5.1.1 Indexicality, Spoken Syntax & Logical Form

Chomsky (1977) assumes that "categories introduced into a base derivation are indexed" (81). While this idea of indexing falls out of vogue and in Chomsky (1995b) it is suggested under the 'condition of inclusiveness' (228) that nothing, including indices, should considered to be introduced into the derivation at the level of syntax, this reconsideration of indices has not been generally incorporated into the working models of most theorists. For example, Baker (2003) notes Chomsky's inclusiveness condition (97) but comments that,

> It does not actually matter to my theory whether these indices are present throughout the computation of a linguistic structure. A legitimate alternative would be that these indices are added at the conceptual-intentional interface, just beyond LF. The substance of my theory can thus be made consistent with Chomsky's (1995) view that indices are not part of the linguistic representation proper. I nevertheless include indices freely in my syntactic representations because I do not know any compelling reason to say they are not there and because it makes the representations more explicit. I leave the exact status of these indices at the different stages of linguistic computation open for further conceptual reflection and empirical research. (97)

In fact, the 'principle of inclusiveness' was proposed for seemingly purely conceptual reasons and, in the context of a rich mechanism for processing syntax by a pencil and paper theorist, there does not seem to be a compelling reason for changing working habits. Deference to this principle are relatively common in the literature (see also footnote 10) but there has been little motivation to change our way of thinking or practice of notation.

---

presume that listener interpretation is a factor in language output.

5.1.2 *Wh*-Word Placement is Limited by Interpretation

In considering the hypothesis that the patterns of apparent *wh*-movement are largely due to whether or not they respect *legibility* conditions (i.e. are they interpretable by a listener?), we must first look at the question of subjacency. Rather than defining subjacency as a stipulative impediment to the formation of certain phrases, I consider it essentially an *anti-reconstruction effect*, whereby the listener is unable to reinterpret it in its canonical position and so cannot interpret it at all.

Wh-words & phrases are particular as a class in being able to be interpreted outside their canonical clause. Nonetheless, when there are more than one *wh*-element in a sentence, superiority must be obeyed where a higher *wh*-element must be interpreted in a 'higher' position if it is not to interfere with the interpretation of a lower *wh*-element (Chomsky (1973), Pesetsky (1987), Richards (2001)); see (68).

(68) a. Who did Catie give what (to) <*who*>

b. What did Catie give who <*what*> (#to) <*who*>

Looking at the problem in terms of legibility conditions (and with consideration for the listener), even if there were indices in a derivation, we must acknowledge that the linguistic string available to the listener is considerably impoverished and free of any potential indices[6]. In this way we can understand the constraint of superiority without stipulation. If we accept the evidence for the copy theory of movement (Chomsky (1995b) & Fox (2002) among others[7]) but instead of considering what is 'left behind' as copies but as elided elements, i.e. ellipsis without indices, we might understand the problem as that of reconstruction for interpretation, just as any ellipsis[8]. In terms of listener interpretation the two are perhaps indistinguishable; the important point is that an elided phrase necessarily would not have an index. When we consider that, even if there were some kind of index on these phonetically null items, for the listener there can be no knowledge of this; if there is any reconstruction or analogy to canonical argument positions, this must be done without benefit of indices. Because any non-speaker interpretation of a sentence must contend with interpretation without overt indices, we can usefully examine the problem in similar ways to that of ellipsis and we might understand the problem as that of reconstruction for interpretation. Thus the grammar is constrained by there being a limited number of possible

[6]This has been suggested independently by (Chomsky, 1995b, 228) as the 'condition of inclusiveness' where he suggests that derivational syntax under the Minimalist Program should not introduce such things as indices.

[7]The argument for the copy theory of movement is based on the observation that movement cannot change the verdict of Condition C (Fox, 2002, 66).

[8]For an example of what I mean by reconstruction of ellipsis, see the ACE example (65) in §4.12, where reconstruction of an elided element can only occur if the antecedent is clear.

antecedents for reconstruction[9]. At some interpretative level there must be a point where co-reference is determined from the index-free syntactic input. This function would fail if there are multiple candidates for reconstruction (as in ellipsis) and so syntax exhibiting subjacency violations would be uninterpretable at some level.

5.1.3 Subjacency and Interpretation

In the example below, the sentence does not parse because 'how' does not seem to modify the phrase with 'wonder'.

(69) * How do you wonder why Catie degreased her bike chain?

This contrasts sharply with a sentence such as "How do you know why Catie degreased her bike chain?" which parses fine because the phrase "how do you know x" is licit. The example in (69) has been described as subjacency violation (Chomsky (1986)); the problem with describing it as such is that it does not provide an explanation beyond being a stipulation of the grammar. At one point Chomsky defined the possibility of crossing one bounding node but not more than one in the following way: "(59) β is n-subjacent to α iff there are fewer than n + 1 barriers for β that exclude α"; while this definition sounds very clear and precise it depends in turn on the definition of 'barrier' which is a notion that, to this day, escapes a definitive description and has led to reformulations in Chomsky (2001), again highly problematic and nowhere as clear as anybody holding onto a purely computational view of syntax would like.

The root of my analysis is in the idea of Relativized Minimality presented in Rizzi (1990), and subsequently adopted into the principle of Locality by Chomsky (1993).

5.1.4 Relativized Minimality or 'Locality' as a systemic necessity

Following Rizzi (1990) and Rizzi (2001), I discuss the extent to which Rizzi's observations called*Relativized Minimality* (henceforth RM) can be understood to constrain a theory of syntax without recourse to many of the stipulations of our current theory. In my view, as much as RM has been adopted within mainstream linguistic theory and reinterpreted as *locality* (particularly in Chomsky (1995b)), its full implications have not been appreciated.

Rizzi's basic observation is that, in (70), one element X cannot refer to another element Y if there is an intervening element of the same category Z that could act as a referent, i.e. an interrupter of the reference between the first two items. Rizzi (2001) states that—in regards to the intervention of Z—intervention could be determined hierarchically in terms of c-command (91).

[9]This is not unlike when chess moves are called out only by their position on the board. If you call Knight to King 4 we know what is going to happen because of the highly constrained circumstances; we do not have to have all the details spelled out.

(70) ...X...Z...Y...

Rizzi describes this in the terms of Government and Binding theory (specifically: government and c-command) but, I argue that this can be interpreted without recourse to such structural notions[10].

That this must be respected whether the movement is presumed to move overtly or covertly is suggested in the following Japanese data. Here is seems that Japanese respects the one *wh*-phrase movement per clause stipulation (discussed in Grohmann (2000)).

(71) a. John-wa [Mary-ga nani-o katta ka dooka] dare-ni tazuneta
John-TOP Mary-NOM what-ACC bought whether who-DAT asked Q
no?

'Who did John ask <*who*> whether Mary bought what?'

b. ??John-wa [Mary-ga nani-o katta ka dooka] Tom-ni
John-TOP Mary-NOM what-ACC bought whether Tom-DAT asked
tazuneta no?
Q
'What did John ask Tom whether Mary bought <*what*>?'

5.2. A Brief Look at Parasitic Gaps and Sluicing

5.2.1 Parasitic Gaps as a Problem for Reconstruction

The principle of failure of interpretation due to unclear reference is very clear when applied to certain parasitic gaps (PGs) constructions. Classic examples of PGs from (Nissenbaum, 2000, 31) are as in (72)

(72) a. Which article did John file _ [before reading_]?

b. ?? Which article did John file *his dissertation* [before reading _]?

c. What movies did Mary claim she liked _ [in order to get you to see _]?

d. ?? What movies did Mary claim she liked *The Godfather* [in order to get you to see _]

e. John's the guy that they said they'll hire _ [if I criticize _ publicly]

f. * John's the guy that they said they'll hire *me* [if I criticize _ publicly]

These sentences are ungrammatical (because uninterpretable) for several reasons, primarily because we have supplied two possible antecedents for an elided element; to indicate the problem, I draw arrows:

[10]cf. also Chomsky's fall 2004 lectures that suggest c-command might fall out from other factors and may in fact be a secondary element

(73) [Which article] did John file [his dissertation] before reading _?

Another reason that the sentence fails is that reconstruction of the *wh*-phrase to its canonical position would fail:

(74) [Which article] did John file [his dissertation]

That is, this is a problem for reconstruction. Although this only touches on the question of PG interpretation, such explanations have largely been ignored in favour of explanations positing null operators (e.g. Chomsky (1986), Nissenbaum (2000)), something without independent verification.

5.2.2 Wh-island effects & Sluicing

While the issue is due more discussion than space permits, sluicing data is also suggestive that manifest forms of syntax matter for the recoverability of co-reference. Ross (1967) observed that sluicing recovers *wh*-word island violations (cf. also Chomsky (1972), Lasnik (2001), (2003), (in press), Merchant (2001)).

(75) a. *Which Marx brother did she say that [a biography of t] is going to be published

b. A biography of one of the Marx brothers is going to be published next week, but I don't know which <*she said that a biography of t is going to be published*>

(76) Weak island (wh-island)

a. ?*Which book did every journalist go out today to find out who was selling t

b. Every journalist went out today to find out who was selling a certain book, but I don't know which <*(book) every journalist went out today to find out who was selling t*>

5.3. A Short Note on the Limits of Reconstruction and Binding Conditions A& C

If the point from which XPs can 'move' is not a trace or exactly a copy we can understand both why Binding condition C violations are not rescued by movement: they are still reconstructed there as ellipsis. This helps explain why an adjunct is not reconstructed–because it was not part of the original serial string containing a given set of argument relations—if it is not one of the arguments of the verb then it was never in the argument domain and can never be reconstructed.

(77) Asymmetric pattern of reconstruction.

a. Which pictures of himself did John sell before Mary had a chance to look at?

b. * Which pictures of himself did Mary sell before John had a chance to look at?

(Nissenbaum, 2000, 40)

If condition A holds at a point where the sentential string is interleaved with any adjuncts and serialized for spoken output, then we can understand this. The phrase 'which picture of himself' needs to be correlated with 'did John sell' and for some reason (for Condition A) Mary comes into it. Therefore a certain *locality* is important. One view here is that C-command is a result of embedding and not necessarily what the grammar is looking for. The question is in fact is Condition A a locality constraint and not a c-command constraint. I.e. can it be satisfied by a local immediate search of either i) same clause (allowing for intervening material) or ii) adjacent local clause that may or may not enter a c-command relation.

From Fox (2002) Adjunct-extraposition bleeds Condition C:

(78) a. I gave him$_i$ a book yesterday that John$_i$ liked

b. ?? I gave him$_i$ a book that John$_i$ liked yesterday

5.4. Tandem Movement

An apparent problem for this proposal is doubly fronted *wh*-words and phrases. Because these are constrained by superiority in languages such as Bulgarian and can only escape superiority effects in languages that allow scrambling, it is possible to interpret the front placement not as two items in need of reconstruction (requiring indices) but as one conjoint placement of two elements. Findings from physical systems such as American Sign Language (ASL) Phonology show that one 'tandem' or co-ordinated action is preferred over two individual actions (Napoli and Wu (2003)). I take this to be a specific example of the common knowledge that the brain has a preference for tandem co-ordinated movements or interpretations over two individuated movements or interpretations. I argue that if doubly fronted *wh*-phrases can be considered as one tandem unit, we can limit *wh*-phrase alternations to one such fronting per domain (e.g. CP, DP).

What is also allowed is 'nested-path' movement where one movement is bounded within an embedded clause (such as a DP) and the other operates within the bounds of a matrix clause or CP—sensitive to the dominant argument structure of the clause (Pesetsky (1982)); these data are also addressed in Richards (2001) under his Principle of Minimal Compliance; see §4.11 for discussion.

Subjacency effects can be explained as the impossibility of interpreting or reconstructing the index-free copy if there is a second possible antecedent. That is, while the metaphor

of 'movement' is useful, it is clearly not the impossibility of pronouncing various ungrammatical examples but the fact that *interpretation* is either difficult or impossible (e.g. ref word salad). Therefore we should not be concerned with barriers to movement but barriers to legibility or interpretation.

5.4.1 Subjacency again

Revisiting the example of subjacency violation from above, we can see the difficulty first adequately described by Rizzi where the relations between the *wh*-phrase and the verb/clause is interrupted and therefore uninterpretable:

(79) How [do you wonder] why [Catie degreased her bike chain?]

That something like subjacency is in fact a problem for interpretation suggests that we have no 'special powers' of interpretation; it seems hard to imagine a more basic problem for general information theory as an interrupter that could be a possible candidate for reference. On the other hand, if subjacency were not a problem, we might have evidence for a biologically 'hardwired' UG. That is, such a system might allow connections similar to the arrows I have indicated as extra structure of the system. In fact, once the arrows are in place on the sentence on the page it is possible (with sizeable difficulty) to see what a phase such as this might mean. That we cannot parse this sentence without such aids does not constitute 'proof' of anything but it is very suggestive that we are prey to the most basic confusions (i.e. like item acting as interrupter of another like item).

Potential landing sites are points at which the *wh*-phrase is not compromised for reconstruction. I.e. if they are filled with another element they are compromised because moving past another *wh*-phrase would create a second antecedent for the point of *wh*-phrase ellipsis, as in (73) above. That is, while most elements are contained at the local level, *wh*-word interpretative reconstruction may be the only barrier to an element escaping the local level.

The following example from Richards (2001) illustrates that superiority must be obeyed even in languages that have covert *wh*-phrase movement and that therefore, on the above assumptions, it is the semantic interface that is unable to parse the legibility of the sentence and that marks the sentence as ungrammatical (and not the scrambling of garebyadnas ('why') over the direct object).

(80) a. Bkrashis- lags- gi gyag garebyadnas gzigs- gnang- pa- red?
Tashi HON ERG yak why buy -HON -PAST -AGR
'Why did Tashi buy a yak?'

b. Bkrashis- lags -gi garebyadnas gyag gzigs- gnang- pa -red
Tashi HON ERG why yak buy HON PAST AGR

(81) a. Bkrashis- lags -gi gagi garebyadnas gzigs- gnang -pa -red
Tashi HON ERG which why buy HON PAST AGR

'Why did Tashi buy what?'

b. *Bkrashis- lags -gi garebyadnas gagi gzigs- gnang- pa -red
Tashi HON ERG why which buy HON PAST AGR

5.4.2 Pesetsky's Path Containment Condition

Pesetsky (1982)'s Path Containment Condition (PCC) is a generalized condition on crossing and nested dependencies (e.g. the Subject condition; ECP) and states that if two (serially determined) paths overlap, one must be contained by the other. The PCC can be illustrated with the the data in (82) (Pesetsky, 1982, 268).

(82) a. What books do you know who to persuade <*who*> to read <*what*>?

b. * Who do you know what books to persuade <*who*> to read <*what*>?

Just as I am suggesting that the ASL data from Napoli and Wu (2003) implies tandem movement respecting Superiority (as in (79)), so too does it suggest a solution for data that moves in mirror-like forms, as in (83).

(83) a. ?Amnesty International is one human rights organization [Oi that I wonder [which atrocityj this official would be easiest [Ok to talk to tk about tj on behalf of ti]]]

b. a.rights organizationi atrocityj officialk tk tj ti

(84) a. *The annexation of East Timor is one atrocity [O_j that I wonder which human rights organization$_i$ this official would be easiest [O_k to talk to t_k about t_j on behalf of t_i]]]

b. atrocity$_j$ rights organization$_i$ official$_k$ t_k t_j t_i

Thus the data that violates the strictness of the mirror form is illicit.

Chapter 6

Conclusion

6.1. The Biological Foundation of Learning?

The debate about whether or not it is possible to *learn* something as arguably complicated as natural language syntax is essentially circular. One theorist may consider it unlikely or impossible for a child to acquire the beginnings of adult syntax in the span of a few years, instead needing some sort of biological hardwiring (e.g. Chomsky (2002); Boeckx and Hornstein (2003) and references within), while another may think of children as clever little scientists who through trial and error deduce what they need to know about their world (e.g. Sampson (1997) in the tradition of Jean Piaget). While these views are defended with different appeals to 'likelihood' or 'scientific reasoning', they remain appeals to our philosophy of how the world must work and as such are statements of opinion[1]. Only once we are able to correlated a theory with evidence from brain physiology can we start to consider these as empirical arguments.

While I cannot help introducing yet another philosophy into the debate, I have attempted to constrain my arguments by assuming the physiological limitations proposed by Mountcastle (1982). Following the lead from Mountcastle's hypothesis, that our search for difference from region to region (hence specialization), may have no significant physiological reality, we must account for linguistic data in the absence of mental 'mechanisms' specialized for language, on the assumption that the neocortex has a common function from region to region. After suggestions by Hawkins (2004), I take this to be at base a system of data storage and recall as well as limited permutation of this data.

This amounts to an investigation into the *biological foundation of learning*. In section §5.1 I outlined some of the constraints on a linguistic system due to limitations on comprehensibility of either or both the speaker or listener due to such things as the absence of

[1] As commented recently by Chomsky in his 2006 Berlin conference address, at the time of his 1974 meeting with Piaget, the "only plausible idea seemed to be that the process is a form of theory construction". If this is true, the heritage of Piaget perhaps shows the philosophical background Chomsky was debating; the philosopher was then debated by the logician. The question remains for the argument to take more regard of the physiology than merely our 'plausible ideas'.

indices. I proposed that the system is greatly constrained by the *failure* to parse certain possible configurations of the syntax. I argue that these are likely limitations on any system of language (human or otherwise) and fully expected without appeals to biological predetermination. That is, they are limits of the grammar due to the *nature of the problem.* This view is in opposition to a theory assuming that principles of language are a direct result of a biologically predetermined system due to our genetic inheritance (i.e. the *innate* view) where it is argued we 'grow' our ability for language. To resolve these views it is important to properly consider the biological basis for learning. While the term 'learning' may be too filled with lay assumptions about intentionality to be interpreted in a fully biological sense, the exact mechanism of learning needs to be addressed if we are not to assume separate functions for learning language, learning faces, learning spacial relations, learning to place the violin etc. That is, can we eventually identify a biologically based reaction to stimuli that is common to the neocortex, or as Piagett would have it, a general function of cognition (Chomsky and Piaget (1980 (1975))?

6.1.1 Considering Child Language as a Form of Language Breakdown

In chapter 4, I suggested that *adult syntax and adjunction can largely be explained if we compare it with child speech errors.* Children are limited by the fact that they have limited control of their articulatory system and may have greater problems with general memory than adults. Nonetheless, as experimentalists such as Steven Crain (e.g. Crain et al. (1996)) have repeatedly demonstrated, they do not so much lack understanding of adult syntax as much as the ability to put it all together. Given these restrictions, the choices they make about which elements in a sentence are most important, we may learn a lot about adult syntactic processing by studying what choices children make when, due to various performance problems, they cannot speak at adult levels. If, as Crain argues, children have 'full competence', then the elements they normally fail to speak might be expected to be dropped because they are not centrally important for the immediate purpose of communication[2]. My hypothesis is that the elements that they speak are not only more important for the discourse but possibly the way in which the child (and consequently the adult) perceives the discourse. That is, they build their arguments not from front to back or back to front but from the verb outwards, proceeding from the verb to the object or indirect object and only later if necessary or possible do they consider the subject.

In this way, instead of using child acquisition data to reenforce findings from adult grammar I suggest that it is possible to use the computational limitation of children to show us how the adult speaker proceeds; this in turn will feed our assumptions about child

[2]While I do not consider the Chomskyan sense of competence/performance to be an accurate dichotomy, Crain's demonstrations of child language ability are valuable demonstrations that there is an important difference between understanding and demonstrated ability.

language acquisition. In essence, I suggest that we take speech errors seriously. Rather than considering child or adult speech 'performance' errors to be unimportant for syntactic and semantic theory, I consider them to inform us about complexities in the computational process that would not be detectable by a theory allowing only 'competence'. The domains of theory and experimental practice can benefit enormously by considering problems in the other.

6.1.2 Extra-systemic Censorship

Another principle I followed, continues with the idea that a theory of language cannot isolate itself from how language is used. As such, I considered what linguistic constraints are introduced by the effort of communicating with a listener. While Chomskyan linguistics has largely disallowed any speculation of speaker intentionality, even a strict nativist might allow that even if the system is fully discrete and self contained, the speaker may likely cease to use certain functions hypothetically allowed by a linguistic component of the brain simply because, when used, comprehension suffers. We must then ask if such self censorship plays a greater role in manifest grammar than normally allowed. It may be the case that we cannot distinguish between certain cases of such conditioned self-censorship and hypothetical 'principles' that the language component might be sensitive to. Nonetheless, we must consider these as their empirical validity is less speculative than the Generative idea of *Principles*.

I used the example of indices to illustrate this point. Indices are generally considered more than just notational shorthand (e.g. Fiengo and May (1994); Chomsky (1995b))[3]. On the assumption that the speaker might add indices to his lexical items in an example of complex discourse as he sends them through his mental representation of syntax, he would then be able to keep track of what he is saying. The listener does not have the benefit of these indices and yet somehow manages to follow most of the permutations of any given sentence. Two questions arise: if the indices are required by the speaker in his mental representation, how is it that the listener does not require them? (Remember that only the crudest indexing is available in the spoken output.) Conversely, if the listener is able to reconstruct the sentence without evidence of indexing, why does the speaker need them in the first place?

That is to say, the syntax has both essentially negative constraints and positive composition factors that act in tandem to limit the possibilities of the syntax. If this division seems overly simple and might leave certain questions unresolved, this is only to be expected. Some theorists claim that unless everything can be explained in the manner I am attempting, then we still need recourse to some biological specification. Few sciences find answers

[3]Rather, if it is no more than our notation, we should beware because our notation is our model and so our shorthand risks becoming part of that.

to their question all in one source. Even on the assumption that we *did* have discretely disposed towards language processing, it would be surprising if the factors I discussed herein did not also condition the grammar.

References

Abels, Klaus. 2003. Successive cyclicity, anti-locality, and adposition stranding. Doctoral Dissertation, University of Connecticut.

Abney, Steven. 1991. Parsing by chunks. In *Principle-based parsing*, ed. Robert Berwick, Steven Abney, and Carol Tenny. Dordrecht: Kluwer Academic Publishers.

Anderson, Stephen R., and David W. Lightfoot. 1999. The human language faculty as an organ. *Annual Review of Physiology* 62:697–722.

Apresjan, J. D., I. Mel'chuk, and A. Zholkovsky. 1969. Semantics and lexicography: Towards a new type of unilingual dictionary. In *Studies in syntax and semantics*, ed. F. Kiefer, 1–33. Reidel: Dordrecht.

Atkinson, Martin. 1992. *Children's syntax: An introduction to principles and parameters theory.* Cambridge, Mass.: Blackwell.

Baker, Mark C. 2003. *Lexical categories.* Cambridge: Cambridge University Press.

Baker, Mark C. 2004. On gerunds and the theory of categories. Ms. Rutgers University.

Beghelli, Filippo, and Tim Stowell. 1997. Distributivity and negation: The syntax of *Each* and *Every*. In *Ways of scope taking*, ed. Anna Szabolcsi, chapter 3, 71–108. Dordrecht: Kluwer Academic Publishers.

Bejar, Susana. 2003. Phi-syntax: A theory of agreement. Doctoral Dissertation, University of Toronto.

Bloom, Paul. 1994. Possible names: The role of syntax-semantics mappings in the acquisition of nominals. *Lingua* 92:297–329.

Boeckx, Cedric, and Norbert Hornstein. 2003. The varying aims of linguistic theory. Ms. University of Maryland.

Boeckx, Cedric, and Massimo Piattelli-Palmarini. in press. Language as a natural object; linguistics as a natural science. *Linguistic Review* .

Borer, Hagit, and Bernhard Rohrbacher. forthcoming. Minding the absent: Arguments for the full competence hypothesis. *Language Acquisition* 0:48.

Borer, Hagit, and Kenneth Wexler. 1987. The maturation of syntax. In *Parameter setting*, ed. Thomas Roeper and Edwin S. Williams, 123–172. Dordrecht, The Netherlands: D. Reidel Publishing Company.

Bruner, J.S., and Leo Postman. 1949. On the perception of incongruity: A paradigm. *Journal of Personality* XVIII:206–223.

Büring, Daniel, and Katharina Hartmann. 1997. Doing the right thing. *The Linguistic Review* 14:1–42.

Bybee, Joan. 1995. Regular morphology and the lexicon. *Language and Cognitive Processes* 10:425–455.

Carey, S. 1978. The child as word learner. In *Linguistic theory and psychological reality*, ed. M. Halle, J. Bresnan, and A. Miller. Cambridge, Mass.: MIT Press.

Carey, S. 1982. Semantic development: The state of the art. In *Language acquisition: The state of the art*, ed. E. Wanner and L.R. Gleitman, 347–389. New York: Cambridge University Press.

Chomsky, Noam. 1955. The logical structure of linguistic theory. Ms. Parts of 1956 revision published in New York: Plenum, 1975, and Chicago: University of Chicago Press, 1985.

Chomsky, Noam. 1965. *Aspects of the theory of syntax.* Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1970. Remarks on nominalization. In *Readings in English transformational grammar*, ed. R. A. Jacobs and Peter S. Rosenbaum, 184–221. Waltham, Mass.: Ginn.

Chomsky, Noam. 1972. *Spheres of influence in the age of imperialism; papers submitted to the bertrand russell centenary symposium, linz, austria, september 11th to 15th, 1972.* Nottingham: Bertrand Russell Peace Foundation for The Spokesman.

Chomsky, Noam. 1973. Conditions on transformations. In *A Festschrift for Morris Halle*, ed. S. Anderson and P. Kiparsky, 232–286. New York: Holt, Rinehart, and Winston.

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

Chomsky, Noam. 1977. On WH-movement. In *Formal syntax*, ed. P. Culicover, T. Wasow, and A. Akmajian. New York, New York: Academic Press.

Chomsky, Noam. 1980. *Rules and representations.* New York: Columbia University Press.

Chomsky, Noam. 1981. *Lectures on government and binding*. Dordrecht, The Netherlands: Foris Publications.

Chomsky, Noam. 1986. *Barriers*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1988. *Language and problems of knowledge: The Managua lectures*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1993. A minimalist program for linguistic theory. In *The view from building 20*, ed. Ken Hale and Jay Keyser, 1–52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 1995a. Bare phrase structure. In *Government binding theory and the minimalist program*, ed. Gert Webelhuth, 383–439. Oxford: Oxford University Press.

Chomsky, Noam. 1995b. *The minimalist program*. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2000a. *The architecture of language*. Oxford, UK: Oxford University Press.

Chomsky, Noam. 2000b. Minimalist inquiries: The framework. In *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, ed. Roger Martin, David Michaels, and Juan Uriagereka, 89–156. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2001. Derivation by phase. In *Ken Hale: A life in linguistics*, ed. Michael Kenstowicz, 1–52. Cambridge, Mass.: MIT Press.

Chomsky, Noam. 2002. *On nature and language*. Cambridge: Cambridge University Press.

Chomsky, Noam. 2004. Three factors in language design. LSA address.

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

Chomsky, Noam, and Howard Lasnik. 1977. Filters and control. *Linguistic Inquiry* 8.

Chomsky, Noam, and Howard Lasnik. 1993. Principles and parameter theory. In *Syntax: an international handbook of contemporary research*, ed. J. Jacobs, Arnim von Stechow, W. Sternefeld, and T. Vennemann, 506–569. Berlin: de Gruyter. Reprinted in Chomsky, N. 1995. The minimalist program, 13–127. Cambridge, Mass.: MIT Press.

Chomsky, Noam, and Jean Piaget. 1980 (1975). Royaumont Symposium proceedings. In *Language and learning: The debate between Jean Piaget and Noam Chomsky*, ed. Massimo Piatelli-Palmarini. Routledge & Kegan Paul.

Cinque, Guglielmo. 1999. *Adverbs and functional heads: A cross-linguistic perspective*. New York: Oxford University Press.

Crain, Stephen, Rosalind Thornton, Carole Boster, Laura Conway, Diane Lilo-Martin, and Elaine Woodams. 1996. Quantification without qualification. *Language Acquisition* 5:83–153.

Croft, William. 1991. *Syntactic categories and grammatical relations*. Chicago: University of Chicago Press.

Déchaine, Rose-Marie. 1993. Predicates across categories. Doctoral Dissertation, University of Massachusetts-Amherst.

Dennett, Daniel C. 1995. *Darwin's dangerous idea*. New York, New York: Simon & Schuster.

Ferdinand, A. 1996. The acquisition of the subject in French. Doctoral Dissertation, HIL/Leiden University.

Feynman, Richard P. 1961. Conservation of momentum. In *Lectures on physics*, volume 1, chapter 10. Basic Books.

Feynman, Richard P., Robert B. Leighton, and Matthew Sands. 1963. Algebra. In *The Feynman lectures on physics*, volume 1, chapter 22. Addison-Wesley.

Fiengo, Robert, and Robert May. 1994. *Indices and identity*. Cambridge, Mass.: MIT Press.

Fisher, Cynthia, D. Geoffrey Hall, Susan Rakowiz, and Lila Gleitman. 1994. When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. *Lingua* 92:333–375.

Fitch, W. Tecumseh. 2005. Dancing to Darwin's tune: a review of *The Singing Neanderthals: The Origins of Music, Language, Mind and Body*, by Steve Mithen. *Nature* 438.

Fox, Danny. 2000. *Economy and semantic interpretation*. Cambridge, Mass.: MIT Press.

Fox, Danny. 2002. Antecedent-contained deletion and the copy theory of movement. *Linguistic Inquiry* 33:63–96.

Freidin, Robert. 1986. Fundamental issues in the theory of binding. In *In studies in the acquisition of anaphora*, ed. Barbara Lust, 151–188. Reidel: Dordrecht.

Geach, Peter. 1962. *Reference and generality*. Ithaca, New York: Cornell University Press.

Gibson, Edward, and Neal J. Pearlmutter. 2000. Distinguishing serial and parallel parsing. *Journal of Psycholinguistic Research* 29:231–240.

Gomez, R.L., and L.-A. Gerken. 1999. Artificial grammar learning by 1 year-olds leads to specific and abstract knowledge. *Cognition* 70:109–135.

Gupta, Anil. 1980. *The logic of common nouns*. New Haven, Conn.: Yale University Press.

Haider, Hubert. 1993. Detached clauses—the later the deeper. Ms. Stuttgart.

Hale, Ken, and Samuel Jay Keyser. 2002. *Prolegomenon to a theory of argument structure*. Cambridge, Mass.: MIT Press.

Hale, Kenneth, and Samuel Jay Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In *The view from building 20*, ed. Kenneth Hale and Samuel Jay Keyser, 53–110. Cambridge, Mass.: MIT Press.

Hauser, M. D., N. Chomsky, and W. T. Fitch. 2002a. The faculty of language: What it is, who has it, and how did it evolve? *Science* 1569–1579.

Hauser, M. D., D. Weiss, and G. F. Marcus. 2002b. Rule learning by cotton-top tamarins. *Cognition* 86:B15–B22.

Hawkins, Jeff with Sandra Blakeslee. 2004. *On intelligence*. New York: Henry Holt.

Heim, Irene, and Angelika Kratzer. 1998. *Semantics in generative grammar*. Malden, Mass.: Blackwell.

Herburger, Elena. 2000. *What counts: focus and quantification*. Cambridge, Mass.: MIT Press.

Hoekstra, Teun, and Nina Hyams. 1998. Aspects of root infinitives. *Lingua* 106:81–112.

Hopper, Paul, and Sandra Thompson. 1984. The discourse basis for lexical categories in generative grammar. *Language* 60:703–752.

Hornstein, Norbert. 1995. *Logical Form: from GB to Minimalism*. Cambridge, Mass.: Basil Blackwell.

Huang, C.-T. James. 1984. On the distribution and reference of empty pronouns. *Linguistic Inquiry* 15:531–574.

Huttenlocher, J., and P. Smiley. 1987. Early word meanings: The case of object names. *Cognitive Psychology* 19:63–89.

Jackendoff, Ray. 2002. *Foundations of language: Brain, meaning, grammar, evolution*. Oxford: Oxford University Press.

Jacob, Fran¸cois. 1976. *The logic of life*. New York: Vintage.

Kayne, Richard S. 1984. 'Unambiguous paths'. In *Connectedness and binary branching*, 129–164. Dordrecht: Foris Publications.

Kayne, Richard S. 1994. *The antisymmetry of syntax*. Cambridge, Mass.: MIT Press.

Kripke, Saul A. 1982. *Wittgenstein on rules and private language: An elementary exposition*. Cambridge, Mass.: Harvard University Press.

Kuhn, Thomas S. 1957. *The Copernican revolution*. Cambridge, Mass.: ***.

Kuhn, Thomas S. 1962. *The structure of scientific revolutions*. Chicago: University of Chicago Press, third edition. Postscript 1970; indexed 1996.

Lakoff, G. 1987. *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago, IL: Chicago University Press.

Lancet, The. 2006. A plea to abandon asthma as a disease concept, *Editorial. The Lancet* 368:705.

Lanier, Jaron. 2000. Artificial stupidity. *Salon.com* October 04.

Larson, Richard. 1988. 'On the double object construction'. *Linguistic Inquiry* 19:335–392.

Larson, Richard, and Gabriel Segal. 1995. *Knowledge of meaning: an introduction to semantic theory*. Cambridge, Mass.: MIT Press.

Lasnik, Howard, and Juan Uriagereka. 2002. On the poverty of the challenge. *The Linguistic Review* 19:147–150. A Review of the 'Poverty of Stimulus Argument', Special Issue.

Lebeaux, David. 1988. Language acquisition and the form of the grammar. Doctoral Dissertation, University of Massachusetts, Amherst.

Legate, Julie Anne, and Charles D. Yang. 2002. Empirical re-assessment of stimulus poverty arguments. *The Linguistic Review* 19:151–162.

Lewis, John D, and Jeffrey L Elman. 2004. Learnability and the statistical structure of language: Poverty of stimulus arguments revisited. In *Proceedings of the 26th Annual Boston University*

Conference on Language Development. UC San Diego & McGill University, and UC San Diego.
Lipski, John M. unpublished. Subject pronoun usage among Spanish dialects. Ms. The Pennsylvania State University.
Marcus, Gary. 2004. *The birth of the mind: How a tiny number of genes creates the complexities of human thought*. New York: Basic Books.
Marcus, Gary F. 2000. Pa bi ku and ga ti ga: Two mechanisms children could use to learn about language and the world. *Current Directions in Psychological Science* 9:145–147.
May, Robert. 1977. The grammar of quantification. Doctoral Dissertation, Massachusetts Institute of Technology.
May, Robert. 1985. *Logical form: Its structure and derivation*. Cambridge, Mass.: MIT Press.
Mel'cuk, Igor. 1988. *Dependency syntax: Theory and practice*. Albany: State University Press of New York.
Mountcastle, Vernon B. 1982. An organizing principle for cerebral function: The unit module and distributed system. In *The mindful brain: Cortical organization and the group-selective theory of higher brain function*, ed. Gerald M. Edelman and Vernon B. Mountcastle. Cambridge, MA: MIT Press.
Napoli, Donna Jo, and Jeff Wu. 2003. Morpheme structure constraints on two-handed signs in American Sign Language: Notions of symmetry. *Sign Language & Linguistics* 6:123–205.
Nissenbaum, Jon. 2000. Investigations of covert phrase movement. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Mass.
Pepperberg, I.M., and J.D. Gordon. 2005. Number comprehension by a Grey Parrot (Psittacus erithacus), including a zero-like concept. *Journal of Comparative Psychology* 119:197–209.
Pepperberg, I.M., M.R. Willner, and L.B. Gravitz. 1997. Development of Piagetian object permanence in a Grey Parrot (Psittacus erithacus). *Journal of Comparative Psychology* 111:63–75.
Pesetsky, David. 1982. Paths and categories. Doctoral Dissertation, Massachusetts Institute of Technology.
Pesetsky, David. 1987. Wh-in-situ: Movement and unselective binding. In *The representation of*, ed. Eric J. Reuland and Alice G. B. ter Meulen. Cambridge, Mass.: MIT Press.
Pfau, Roland. 2000. Features and categories in language production. Doctoral Dissertation, Johann Wolfgang Goethe-Universität zu Frankfurt am Main.
Piatelli-Palmarini, Massimo, and Juan Uriagereka. 2005. The immune syntax: The evolution of the language virus. In *Universals and variations in biolinguistics*, ed. L. Jenkins. London: Elsevier.
Pierce, Amy. 1992. *Language acquisition and syntactic theory: a comparative analysis of French and English child grammars*. Dordrecht: Kluwer.
Pinker, Steven. 1984. *Language learnability and language learning*. Cambridge, Mass.: Harvard University Press.
Pinker, Steven. 1994. *The language instinct: How the mind creates language*. New York, New York: Harper Collins.
Pinker, Steven. 1995. Language acquisition. In *An invitation to cognitive science: Language*, ed. Lila R. Gleitman and Mark Liberman, volume 1, chapter 6, 135–182. Cambridge, Mass.: MIT Press.
Pinker, Steven, and Ray Jackendoff. 2005. The faculty of language: What's special about it? *Cognition* .
Poeppel, David, and Kenneth Wexler. 1993. The full competence hypothesis of clause structure in early German. *Language* 69:1–33.
Pollock, Jean-Yves. 1989. Verb movement, UG and the structure of IP. *Linguistic Inquiry* 20:365–424.
Premack, D. 1990. Words: What are they, and do animals have them? *Cognition* 37:197–212.
Pullum, Geoffrey, and Barbara Scholz. 2002. Empirical assessment of stimulus poverty arguments. *The Linguistic Review* 19:8–50.
Richards, Norvin. 2001. *Movement in language: interactions and architectures*. Oxford: Oxford University Press.
Rivero, Maria Luisa. 1994. Clause structure and V-movement in the languages of the Balkans. *Natural Language and Linguistic Theory* 12:63–120.

Rizzi, Luigi. 1990. *Relativized minimality*. Cambridge, Mass.: MIT Press.
Rizzi, Luigi. 1994. Some notes on linguistic theory and language development: the case of root infinitives. *Language Acquisition* 3:371–393.
Rizzi, Luigi. 1997. 'The fine structure of the left periphery'. In *Elements of grammar*, ed. Liliane Haegeman, 281–337. Dordrecht: Kluwer.
Rizzi, Luigi. 2000. *Comparative syntax and language acquisition*. New York, New York: Routledge.
Rizzi, Luigi. 2001. Relativized minimality effects. In *The handbook of contemporary syntactic theory*, ed. Mark Baltin and Chris Collins, 89–110. Oxford: Blackwell Publishers.
Roeper, Tom, and Bernhard Rohrbacher. 1994. Null subjects in early child english and the theory of economy of projection. *University of Pennsylvania Working Papers in Linguistics* 2:83–119.
Ross, John. 1967. Constraints on variables in syntax. Doctoral Dissertation, Massachusetts Institute of Technology.
Sag, Ivan. 1976. Deletion and logical form. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, Mass.
Sampson, Geoffrey. 1989. Language acquisition: growth or learning? *Philosophical Papers* 18:203–240.
Sampson, Geoffrey. 1997. *Educating Eve: the language instinct debate*. New York: Cassel Academic Publishers.
Saussure, Ferdinand de. 1916. *Cours de linguistique générale*. France: Grande Bibliotheque Payot.
Simon, Herbert. 1962. The architecture of complexity. In *Proceedings of the American Philosophical Society*, volume 106, 467–482. Reprinted in H.A. Simon, *The Sciences of the Artificial* MIT Press, 1969.
Skinner, B.F. 1957. *Verbal behavior*. New York: Appleton-Century-Crofts.
Stuurman, Frits. 1985. *Phrase structure theory in generative grammar*. Dordrecht: Foris.
Tenny, Carol L., and Peggy Speas. 2004. The interaction of clausal syntax, discourse roles, and information structure in questions. Talk presented at Workshop on Syntax, Semantics and Pragmatics of Questions. European Society for Logic, Language, and Information (ESSLI), Nancy, France., Carnegie Mellon University and University of Massachusetts at Amherst.
Travis, Lisa. 1984. Parameters and effects of word order variation. Doctoral Dissertation, Massachusetts Institute of Technology, Cambridge, MA.
Uriagereka, Juan. 2002. *Derivations: Exploring the dynamics of syntax*. London: Routledge.
Wexler, Ken. 1994. Finiteness and head movement in early child grammars. In *Verb movement*, ed. David Lightfoot and Norbert Hornstein, 305–350. Cambridge: Cambridge University Press.
Wiggins, David. 1980. *Sameness and substance*. Cambridge, Mass.: Harvard University Press.
Wijnen, F. 1996. Temporal reference and eventivity in root infinitives. *MIT Occasional Papers in Linguistics* 12:1–25.
Wittgenstein, Ludwig. 1958. *Philosophische Untersuchungen*. London: Blackwell.
Yang, Charles. 2002. *Knowledge and learning in natural language*. Oxford: Oxford University Press.