

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RECURSION, SALTATION, BUT NOT COMMUNICATION?
A CRITIQUE OF CHOMSKY'S LANGUAGE EVOLUTION THEORY

THESIS
PRESENTED
AS PARTIAL REQUIREMENT
TO THE MASTERS IN PHILOSOPHY

BY
ERIC MUSZYNSKI

APRIL 2015

UNIVERSITÉ DU QUÉBEC À MONTRÉAL
Service des bibliothèques

Avertissement

La diffusion de ce mémoire se fait dans le respect des droits de son auteur, qui a signé le formulaire *Autorisation de reproduire et de diffuser un travail de recherche de cycles supérieurs* (SDU-522 – Rév.01-2006). Cette autorisation stipule que «conformément à l'article 11 du Règlement no 8 des études de cycles supérieurs, [l'auteur] concède à l'Université du Québec à Montréal une licence non exclusive d'utilisation et de publication de la totalité ou d'une partie importante de [son] travail de recherche pour des fins pédagogiques et non commerciales. Plus précisément, [l'auteur] autorise l'Université du Québec à Montréal à reproduire, diffuser, prêter, distribuer ou vendre des copies de [son] travail de recherche à des fins non commerciales sur quelque support que ce soit, y compris l'Internet. Cette licence et cette autorisation n'entraînent pas une renonciation de [la] part [de l'auteur] à [ses] droits moraux ni à [ses] droits de propriété intellectuelle. Sauf entente contraire, [l'auteur] conserve la liberté de diffuser et de commercialiser ou non ce travail dont [il] possède un exemplaire.»

UNIVERSITÉ DU QUÉBEC À MONTRÉAL

RÉCURSION, SALTATION, MAIS SANS COMMUNICATION ?

CRITIQUE DE LA THÉORIE CHOMSKYENNE DE
L'ÉVOLUTION DU LANGAGE

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN PHILOSOPHIE

PAR

ERIC MUSZYNSKI

AVRIL 2015

REMERCIEMENTS

Je tiens à remercier mon directeur de mémoire, Luc Faucher, ainsi que les membres du jury, Pierre Poirier et Jean-Guy Meunier, pour des commentaires pertinents et suggestions appréciées tout au long de l'élaboration de ce travail.

Merci également au Conseil de recherches en sciences humaines (CRSH) et au Fonds de recherche société et culture Québec (FRQSC) pour le soutien financier durant cette recherche.

Finalement, merci à Charles Reiss, linguiste à l'université Concordia, pour m'avoir introduit au monde fascinant de la linguistique de Chomsky, Clare Fawcett, anthropologue à l'université St-Francis Xavier, pour son aide à plusieurs niveaux, et Katherine McClintock pour la révision de l'anglais; toutes les erreurs demeurent, bien sûr, miennes.

TABLE OF CONTENTS

REMERCIEMENTS	i
TABLE OF CONTENTS	ii
LIST OF FIGURES	
RÉSUMÉ	
ABSTRACT	
INTRODUCTION	1
CHAPTER I	9
THE "RECURSION-ONLY" HYPOTHESIS	
1.1. Concepts of evolution.....	9
1.1.1. Adaptation.....	10
1.1.2. Exaptation	10
1.1.3. Gradualism versus saltationism	11
1.2. FLN/FLB distinction	14
1.2.1. The distinction defined.....	14
1.2.2. FLB and its contents	16
1.3. FLN and the saltationist argument	19
1.3.1. The so-called "recursion-only" hypothesis	19
1.3.2. Critiques by Pinker and Jackendoff	19
1.3.3. Language's appearance through saltation	21
1.4. Recursion defined.....	22
1.4.1. Recursion – definition 1: embedding	23
1.4.2. Recursion – definition 2: cyclical	24
1.4.3. Effects of recursion	25

1.5. Critiques to recursion-only	26
1.5.1. Karlsson: "True" recursion is rare.....	27
1.5.2. Critique of Karlsson	28
1.5.3. Everett: Pirahã.....	30
1.5.4. The "official" answer and its pitfalls	31
1.6. Recursion as a property	33
1.6.1. Recursion instantiated	34
1.6.2. The recursive property and the interfaces	36
1.7. Conclusion.....	37
CHAPTER II.....	39
THE MINIMALIST PROGRAM	
2.1. Beyond the "recursion-only" hypothesis.....	39
2.2. The Minimalist Program as field of research	40
2.2.1. Brief history	40
2.2.2. Aims	42
2.3. Building blocks of the MP	44
2.3.1. Merge	44
2.3.2. Lexical items	46
2.4. Application of Merge and lexical features	48
2.4.1. Asymmetry of interfaces	49
2.4.2. Additional complexities in syntax.....	51
2.5. The place of Merge in the faculty of language.....	53
2.5.1. Computational load	53
2.5.2. Merge and recursion.....	53

2.6. The broader scope of Merge.....	54
2.6.1. In human cognition	55
2.6.2. In non-human cognition	56
2.7. Conclusion.....	59
CHAPTER III	60
MERGE AND SALTATIONISM	
3.1. Chomsky's Minimalism and language evolution	60
3.2. Premises.....	62
3.2.1. Setting the stage	62
3.2.2. Human revolution and "Out of Africa"	63
3.2.3. Species-wide language faculty	64
3.2.4. Language not for communication	64
3.2.5. The atoms of computation.....	66
3.3. Chomsky's conclusions	67
3.3.1. Merge as keystone.....	67
3.3.2. Communication's non-adaptationist origins	69
3.3.3. No proto-language.....	70
3.3.4. Chronology of language evolution.....	70
3.4. Strengths of the theory	71
3.5. Critiques of premises.....	73
3.5.1. The Minimalist Program at center stage	74
3.5.2. Human revolution.....	74
3.5.3. Species-wide language faculty	76
3.5.4. Language not for communication	76

3.5.5.	The atoms of computation.....	80
3.5.6.	Premises: summary	80
3.6.	Chomsky's conclusions revisited	81
3.6.1.	Merge as keystone.....	83
3.6.2.	Communication's non-adaptationist origins	84
3.6.3.	No proto-language.....	88
3.7.	Chronology revisited	89
3.7.1.	The case of the missing lexical items.....	90
3.7.2.	Boeckx: saltationism through lexical items	91
3.7.3.	Jackendoff: the MP precludes gradualism	92
3.7.4.	Clark: a moderate approach	94
3.7.5.	The MP does not warrant saltationism.....	94
3.7.6.	An alternative scenario.....	95
3.8.	Conclusion.....	99
CONCLUSION		101
BIBLIOGRAPHY		104

LIST OF FIGURES

FIGURE 1 - Example of a derivation through four applications of Merge.....	48
--	----

RÉSUMÉ

Noam Chomsky défend depuis longtemps une conception de la faculté du langage voulant que celle-ci ait des assises biologiques. Néanmoins, son implication dans les débats entourant l'évolution de la faculté du langage est chose récente. Il défend la position que le langage aurait évolué par une saltation : en d'autres mots, il serait apparu grâce à un seul changement génétique, et n'aurait pas été le fruit d'une adaptation pour la communication. Le travail qui suit examine et critique les arguments mis de l'avant par Chomsky pour supporter cette idée depuis son entrée dans le débat. Sa première contribution explicite à ce sujet (Hauser, Chomsky et Fitch, 2002) propose que la seule caractéristique du langage qui lui est propre est la récursion. L'article propose que, puisqu'on retrouve toutes les autres propriétés du langage soit dans d'autres aspects de la cognition humaine soit dans d'autres espèces, il s'ensuit que la récursion est le seul élément qui a dû évoluer spécifiquement pour le langage. Un examen approfondi de cette proposition soulève deux problèmes majeurs : la récursion n'est jamais définie, et l'algorithme qui doit instancier cette récursion n'est jamais spécifié. Chomsky a subséquemment publié des articles (Chomsky, 2010; Berwick & Chomsky, 2011) où ces problèmes sont résolus en faisant appel aux découvertes faites dans le cadre du Programme Minimaliste, sa théorie linguistique la plus récente. Après un bref aperçu de ce Programme, les prémisses et les conclusions de son travail récent à propos de l'évolution du langage sont examinées pour montrer comment elles appuient son approche saltationiste. Une telle étude révèle où se situent les failles potentielles de la proposition, ce qui permet de démontrer que le scénario saltationiste n'est pas justifié adéquatement, même sans sortir du paradigme du Programme Minimaliste. Finalement, je présente un scénario gradualiste de l'évolution du langage utilisant les mêmes mécanismes que ceux proposés par Chomsky, montrant ainsi qu'une telle approche est plus plausible que l'alternative saltationiste.

Mots-clés : évolution du langage, Programme Minimaliste, biolinguistique, saltationisme

ABSTRACT

Noam Chomsky has long argued for the biological basis of our species' language faculty, but has only recently involved himself in debates surrounding language evolution. His position is that language evolved through a saltation: in other words, language appeared in a single evolutionary step, and was not the result of adaptation for communication. This paper follows and criticizes the arguments advanced by Chomsky to defend this claim since his entry into the debate. His views were first made explicit in Hauser, Chomsky & Fitch (2002), wherein it is proposed that the sole characteristic of language differentiating it from other cognitive capacities in humans and animals is recursion. This implies that it is the only property that needed to have evolved specifically for language. A thorough review and critique of the article shows that it suffers from two major setbacks: a lack of definition of recursion, and the fact that the algorithm instantiating the recursion in question remains unspecified. Chomsky later published additional articles which resolved these issues (Chomsky, 2010; Berwick & Chomsky, 2011), drawing on findings in the Minimalist Program, his latest linguistic theory. After a brief overview of the Program, the premises and conclusions of his most recent work on language evolution are scrutinized, which defend a saltationist account of the evolution of the language faculty. This also reveals certain flaws in the argument, allowing me to argue that the saltationist scenario is unwarranted, even when one takes at face value the findings within the Minimalist Program. I conclude by proposing a gradualist scenario to demonstrate that it is more plausible than the saltationist account.

Keywords: language evolution, Minimalist Program, biolinguistics, saltationism

INTRODUCTION

Study into the evolution of the human language faculty logically ought to draw on findings in contemporary linguistics. Starting in the mid-twentieth century, thanks in no small measure to the work of Noam Chomsky, the field of linguistics changed its focus from the structural regularities found in languages to the underlying cognitive mechanisms which allow language comprehension and production. Chomsky argued in a famous debate with Piaget (Piattelli-Palmarini, 1980) that this capacity should be understood as a "language organ" assumed to be common to all human beings. The fact that the human species is endowed with a biological mechanisms for human language should not be controversial: it is evident that only humans communicate using human words, syntax, phonology and the whole gamut of mechanisms that humans use to produce language. This is not to say that other species do not communicate, nor even that they do not have their own "language"; it implies merely that humans have their own species-specific language, which, taken as a whole, is inaccessible to other species. No matter how much you may talk to your dog or your plants, they will never answer using a sentence in English (or Hindi or Inuktitut...).

The present investigation generally uncritically takes for granted much of Chomsky's framework regarding the study of language, but, as will be clear, does so in an effort to reveal certain flaws when it is applied to language evolution. When speaking of "language evolution" in this context, it is important to understand that this research is specifically about the biological evolution of the capacity for language as found in the human species. From a "Chomskyan" perspective, the interest of the research is not to study a general capacity for communication, including as it would "body language", social norms, prescriptive grammar, and perhaps even any information-conveying mechanism in any biological entity. Such a delineation would be far too broad, not to

mention that to systematically address all the relevant mechanisms in a uniform manner may be (near) impossible (Chomsky, 2011). The cultural evolution of specific languages is also set aside; different cultures have produced a wide variety of languages, and disciplines such as historical linguistics have much to say about the spread, interactions, and development of particular languages. Culture may moreover play an important role in providing feedback to biological mechanisms, in a process called gene-culture co-evolution (Laland & Brown, 2011, chap. 7), or niche-construction theory (Bickerton, 2009a, pp. 98-105). Chomsky however has maintained a strict separation between the biological aspects of language—those which concern him—, and those stemming from culture, or the environment more generally (Chomsky, 2005). His research therefore investigates the faculty of language, and not the plurality of languages, or a language in particular. Of course this very restrictive delineation of the object of study is by no means adopted by all researchers of language evolution, as will be shown shortly. But for the present research, I propose to provisionally adopt Chomsky's approach to better show its internal inconsistencies. With this in mind, I will be using the expressions “language evolution” and “evolution of language” to refer to the evolution of the biological capacity of language in the human species.

Interestingly, despite Chomsky's emphasis on ties between language and biology, for the greater part of his career he seemed reluctant to delve into the issue of language evolution. Until 2002, most of his comments on the matter were noncommittal statements to the effect that it is “hard to imagine a course of selection that could have resulted in language” (quoted in Kenneally, 2007, p. 39)¹, with the implied

¹ This quotation is curiously not attributed to a specific publication in Kenneally's book, though it may come from an interview conducted by Kenneally. See Botha (1998) for an in-depth review of quotes by Chomsky pre-2002 about his view of language evolution. All quotes are qualified and no specific argument or conclusion is clearly present, although the general sentiment seems to be that expressed in Kenneally's passage.

conclusion that there is little of interest to say about language evolution. As has often been pointed out, the study of the evolution of the faculty of language is notoriously difficult because of a general lack of evidence. Language does not leave fossils, and evidence must therefore be inferred from our knowledge of the contemporary language faculty, reverse engineering of linguistic capacities, comparative studies of proto-linguistic capacities found in other species, as well as extrapolations from archaeological findings of cultural artefacts. But it is very difficult to make accurate inferences from this type of research and such an endeavour can easily lead to “Just-So stories” about language evolution: stories that sound nice, maybe even plausible, but are not truly supported by any hard evidence. It has become almost hackneyed for works such as this one to mention the 1866 ban on publications about the evolution of language by the *Société Linguistique de Paris* and the *Philological Society of London*, typically thought to be in reaction to the proliferation of unfounded and indeed unfalsifiable speculations about the origins of language². That sentiment is still present to this day, notably echoed in Lewontin’s (1998) pessimistic argument that there is much, if not all, of the evolution of cognition which we will never be able to explain.

Nevertheless, many researchers from diverse fields have disregarded such pessimistic arguments and have developed various approaches to the evolution of language. Earlier in the twentieth century, Leroi-Gourhan (1964) argued from neurophysiological and archaeological evidence that bipedalism had the significant effect of freeing the hands and face for fine motor use. The development of tool-making and speaking would have developed in parallel, reinforcing one another. More recently,

² But cf. Cohen’s (2013) “Historical, Darwinian, and current perspectives on the origin(s) of language”, Lefebvre, Claire, Bernard Comrie and Henri Cohen (eds.), *New Perspectives on the Origins of Language*. xvi, 582 pp. (pp. 3–30) wherein he argues that at least in the case of the *Société Linguistique de Paris*, the ban may have been the result of political and sectorial pressures rather than considerations regarding the value of such work.

Philip Lieberman, a cognitive scientist with a background in phonetics, has studied the motor control and cognitive functions involved in speech production, including data regarding the human species' vocal tract length. According to his research, the length is unique in allowing the production of our human phonemic repertoire (Lieberman, Klatt, & Wilson, 1969). He also looked into the interaction of various parts of the cerebral cortex, arguing that the idea of a "language organ", or module, is erroneous. Our capacity for language is instead the result of interactions of various parts of the brain also implicated in other capacities, which were mostly already present in our primate ancestors (cf. Lieberman 2002). Along similar lines, Sue Savage-Rumbaugh and her collaborators (Savage-Rumbaugh & Lewin: 1994; Savage-Rumbaugh, Shanker & Taylor: 1996, 1998) work on primate research, teaching human language to bonobos in an attempt to demonstrate the similarities between humans and one of the most closely related species to humans. If apes could be shown to learn language in the same way we do, it would suggest continuity in cognitive capacities between humans and apes, implying that human linguistic capacities are not unique to humans, merely more developed versions of primate cognition. She claims that some of her bonobos have successfully acquired human language. Her research is however contradicted by Terrace's (1979) work with the (then) famous Nim Chimpsky, a chimpanzee raised in a human family and taught sign language; despite learning some vocabulary, Nim never spontaneously created complete grammatical sentences.

One of the most influential articles in the field of linguistics was Steven Pinker and Paul Bloom's (1990) "Natural Language and Natural Selection", in which they explicitly argued against what they perceived to be Chomsky's position on language evolution. In the article they argue that "language is no different from other complex abilities such as echolocation or stereopsis, and that the only way to explain the origin of such abilities is through the theory of natural selection" (1990, p. 3). Pinker & Bloom further hypothesize that language would have evolved for the purpose of

communication, with gradual changes being the fruit of natural selection for this capacity (ibid, p. 14). Others have proposed specific capacities or selection pressures which would have accounted for the origin of language. For instance Bickerton (2009a) proposes that displacement—the capacity to refer to things outside our immediate perception—would have been the first ingredient for the development of language. Dunbar (1996) hypothesises that language replaced grooming as a form of social bonding in our ancestors when our communities became too large to allow for sufficient grooming time between each individual. Despite these and many other proposals regarding the evolution of language, none of these arguments seemed to hold sway over Chomsky, who stayed surprisingly mute on such matters.

All this changed when in 2002, Chomsky coauthored an article with two evolutionary biologists, Marc D. Hauser and Tecumseh Fitch, for *Science*: “The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?” In this paper Chomsky’s views on language evolution were finally made explicit. Meant to facilitate interdisciplinary communication, the article proposed a distinction between the Faculty of Language – Broad (FLB) and the Faculty of Language – Narrow (FLN). Whereas FLN contains all the biological components that are unique to human language and to the human species, FLB contains FLN as well as the components which contribute to the language capacity, but are also found in other parts of cognition or in other species. The authors further hypothesized that the only thing that satisfies the uniqueness condition is “recursion”—a term which was unfortunately never explicitly defined in the article. This, it is implied, justifies a saltationist approach to the evolution of language. As will be explained in more detail in chapter I, in biology a saltation (from the latin “saltus”, leap) is a sudden, atypical change in a species over a single generation, as opposed to the more common gradual change generally expected through slight variation and natural selection. The argument advanced by Hauser *et al.* (2002) to justify this approach is that the parts of language which are not unique to language would have already been in place (since they are

present in other species) allowing recursion—ostensibly a very simple mechanism—to appear and act as a keystone to all other components, effectively uniting all the parts necessary for a fully functioning language capacity in one step. Furthermore, according to the authors, if whatever satisfies the uniqueness condition is indeed that restricted, it “has the interesting effect of nullifying the argument from design” (2002, p. 1573). In other words, since such a simple mechanism simply cannot appear in more than a single step, it could not be the result of successive steps of natural selection, shaping it to its use in language. The hypothesis that it is not the fruit of adaptation also suggests that it may be the result of non-evolution-related constraints that may be present in neurological architecture.

This article is critically assessed in chapter I, where I review and clarify the arguments sketched above, as well as detail some of the criticisms it has drawn. After a brief overview of central concepts to evolutionary biology, the FLB/FLN distinction is examined. I argue that rather than facilitating cross-disciplinary communication, it may have caused more confusion than anything else. I then turn to the sole component hypothesized to be in FLN: recursion. Various definitions of recursion are explored, showing that since no definition is given in the article, what the authors meant is far from obvious. After a review of two previous critiques of the recursion-as-FLN hypothesis, I propose a critique of my own which, unlike the two previous ones, has the advantage of not relying on a change of theoretical framework or on empirical data, but simply the conceptual clarification of the term “recursion” and its place within FLN. This approach reveals that if no specific recursive algorithm is specified for FLN, Hauser, Chomsky and Fitch’s hypothesis is doomed to remain inconclusive and thus contributes very little to the debate.

After 2002, Chomsky penned additional articles, developing his view of language evolution, emphasizing his saltationist and non-adaptationist account. In these articles, he bases much of his argument on findings in the Minimalist Program, the

most recent generative linguistics paradigm for the study of language. The Minimalist Program attempts to reconcile linguistics and biology, “forc[ing] linguists to reformulate previous findings in terms of elementary units, operations and interface conditions” (Di Scullio & Boeckx, 2011, p. 4) in the hopes that such simplified mechanisms would more easily be found through biology. As will be explained in greater detail in chapter II, the Minimalist Program postulates a single syntactic process “Merge”, which joins two lexical units (words or previous outputs of Merge), ensuring grammaticality thanks to lexical features on the units, in a process called “feature-checking”. Simplifying to the utmost, lexical units have features, for instance the feature [+f], which need to be matched by another unit in the Merge operation, in this case a lexical unit with the feature [-f]; successive applications of Merge continue until all features are checked, resulting in a well-formed sentence (Berwick, 2011, p. 87). In Chomsky’s writings, Merge has come to replace recursion as the uniquely human linguistic capacity, rebutting some of the critiques levelled at the Hauser et al. (2002) article. It will however become clear that his approach offloads the computational complexity onto the lexical features, seemingly allowing Merge to be the simple algorithm needed for the saltationist argument. As will be shown, Chomsky’s reliance on the Minimalist program effectively resolves the critiques levelled at him in chapter I, but opens him up to other critiques formulated in chapter III.

Researchers have rejected Chomsky’s ideas regarding the evolution of language on many fronts. Needless to say, his saltationist and non-adaptationist conclusions are often viewed unfavourably, but few have been able to point out inherent problems within Chomsky’s account. Some will invoke previously formulated alternative hypotheses, often developed by researchers outside the field of linguistics, approaching the question from another angle (see for instance the research mentioned earlier by Lieberman *et al.*, 1969, Lieberman 2002; Dunbar, 1996; Savage-Rumbaugh & Lewin, 1994; Savage-Rumbaugh *et al.*, 1996, 1998). Others will accept Chomsky’s

approach to linguistics in general, but argue that the empirical evidence does not—or likely will not in the future—support his theories (cf. Pinker & Jackendoff, 2005; Jackendoff & Pinker, 2005; Bickerton 2009b, 2009c). In chapter III, I propose instead to argue from within Chomsky's framework and reveal through careful examination that his theory contains certain inconsistencies. I argue that Chomsky's new approach fails because it must disregard the use of lexical features in order to maintain its saltationism. I explore in greater detail the explicit saltationist arguments as found in Chomsky (2010) and Berwick & Chomsky (2011) to show that the appearance of lexical units is assumed to have happened before Merge. As has been remarked explicitly or implicitly in a handful of articles (cf. Bickerton, 2009b; Jackendoff, 2010; Boeckx 2011; Clark, 2013), but not fully fleshed out into a strong argument, the evolution of lexical items and features is a major stumbling block for Chomsky's saltationist argument. I argue that it is more plausible for Merge to have developed much earlier, with lexical features arriving later and developing gradually, eventually allowing for the linguistic competence we now have.

Through careful examination of the premises and arguments advanced by Chomsky and his co-authors, I will demonstrate that Chomsky's arguments for language's development through saltation fail to capture the implications of the various interacting mechanisms of language that he himself proposes. In so doing, I show that his theory is at best incomplete, and at worst incoherent. It furthermore suggests that if the Minimalist Program is to have a strong impact on language evolution debates, more research is required into the relation between words, concepts, and lexical features, a path which has the potential to lead to a more plausible, intuitive, and gradualist account of language evolution.

CHAPTER I

THE "RECURSION-ONLY" HYPOTHESIS

1.1. Concepts of evolution

Before delving into language evolution proper, a few of the central concepts of evolutionary theory must be defined in order to properly situate the discussion in Hauser, Chomsky and Fitch (2002; henceforth HCF) and Chomsky's writings. Interestingly however, many of the discussions on language evolution that will be surveyed in this paper do not engage in significant ways with much of the work in biology or the conceptual distinctions proposed in the philosophy of biology³. This is undoubtedly a lacuna, as many arguments rely on sometimes poorly understood principles of evolutionary biology, especially in the articles Chomsky has published after his work with Hauser and Fitch (see chapter III). In any case, no matter the accuracy with which it is deployed by Chomsky or others, background knowledge of central and relevant themes in evolution is necessary to properly understand the issues and arguments pertaining to language evolution.

Because this thesis focuses specifically on Chomsky's theoretical framework, I will mostly be relying on concepts he himself has used, namely exaptation and saltationism, which are, for the most part, sufficient for the purposes of my critiques.

³ Other sources of course do approach the issue from the biological perspective, but they generally do not engage directly with the work done by Chomsky, and therefore rarely meet on common ground.

1.1.1. Adaptation

Evolution is “the theory of the change of organic species over time” (Sloan, 2010). Lewontin has characterized our contemporary understanding of evolution through three necessary building blocks of evolution: (1) variations in morphologies, physiologies and behaviour (“phenotypic variation”), which leads to (2) varying rates of survival and reproduction (“differential fitness”), which must also be (3) heritable (“fitness is heritable”) (Lewontin, 1970, p. 1). Acting upon these mechanisms is natural selection, famously proposed by Darwin (1859), which will tend to eliminate those organisms with the least fitness, leading to change in the population over time. The typical assumption within the theory of evolution is that traits which are present in a species are adaptations. (However this assumption is sometimes faulty, as will be shown in the next section.) Adaptations are traits that have been favoured by natural selection (Laland & Brown, 2011, p. 92), meaning that they are adapted to a specific environment, or niche.

1.1.2. Exaptation

Yet not all traits are selected for. Williams (1966), an influential evolutionary biologist, stated that “adaptation is an onerous concept that should be used only where it is really necessary” (p. 4). In the same vein, Gould and Lewontin (1979) in their well-known article “The Spandrels of San Marco” argue that evolutionary biologists too often assume that the only explanation for any given trait is that it is an adaptation. Many factors other than adaptation can explain the presence of a given trait, such as architectural constraints (p.15), by-products of other adaptations, otherwise known as spandrels (pp. 2-3), or genetic drift (p.10) among others. Gould

and Vrba (1982) coined the term “exaptation” to refer to a trait which contributes to the fitness of an organism without having been selected for that specific function.

An exaptation is a trait that is currently useful for a given function, but whose historical development is not (at least not exclusively) the result of selection for that function. In the words of Gould and Vrba: “such characters, evolved for other usages (or no function at all), and later “coopted” for their current role, [are] called exaptations” (1982, p. 6). For example, the development of feathers in birds apparently happened before flight, implying that feathers could not have developed as an adaptation for flight, despite their obvious contemporary importance with respect to that function. Gould and Vrba mention a few hypotheses regarding their origin, such as the possibility that they developed for insulation, only later being exapted for flight (ibid., p. 7). Note that this does not exclude the fact that feathers or any other exaptation could have subsequently evolved for that new function, only that at some point previously in the trait’s development, it was *not* selected for the current function. Exaptations and adaptations are thus not mutually exclusive.

1.1.3. Gradualism versus saltationism

Another typical assumption within the theory of evolution is that changes within a population happen gradually. Darwin believed that “*Natura non facit saltum*”, which is to say that evolutionary change must happen in small successive steps rather than great leaps. He even stated that “if it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down.” (Darwin 1859, p. 189) For Darwin, and indeed in the mind of many, evolution must be the result of slight variations in successive generations.

But Darwin's gradualism came under fire by saltationists as soon as it was proposed. Even his supporter Thomas Huxley, "Darwin's bulldog", warned him in a letter that he had loaded his work "with an unnecessary difficulty in adopting '*Natura non facit saltum*' so unreservedly" (Huxley, 1859). The debate continues into the twentieth century, with Richard Goldschmidt (1940, paraphrased in Gould 1980) proposing that every so often, a great mutation will produce a "hopeful monster", which may lead to a rapid, discontinuous change from one generation to the next. This idea is later defended by Stephen Jay Gould, who explains Goldschmidt's idea in the following way:

new species arise abruptly by discontinuous variation... every once in a while a macromutation might, by sheer good fortune, adapt an organism to a new mode of life, a "hopeful monster" in his terminology. Macroevolution proceeds by the rare success of these hopeful monsters, not by an accumulation of small changes within populations. (Gould, 1980, p. 24)

Gould then goes on to illustrate the success of such an approach through examples of species which have features that purportedly allow for no intermediate gradual steps. For instance, a certain genera of boid snakes have a movable joint in the center of the upper jaw, which seems to have no possibility for gradual adaptation: "How can a jawbone be half broken?" (Gould, 1980, p. 26) This is precisely the type of argument invoked by HCF and later Chomsky for their saltationist claims.

Saltationism is often associated with speciation and punctuated equilibrium, but that association is not fully warranted in either case. As testified by the quotation in the previous paragraph, saltations can account for speciation, but speciation can also happen through gradual evolution, for instance in the case where a given population is segregated from another for long enough that the gradual changes amount to a new species. Punctuated equilibrium is a theory that proposes that those gradual changes that lead to speciation typically happen in bursts (on a geological timescale) rather than slow accumulation of slight changes (Gould & Elderedge, 1977). Although the

changes leading from one species to the next are thought to be rapid, they are nevertheless gradual and not saltational (Gould, 2002). Saltations thus do not equate with speciation, although they are one way in which it may happen.

The discontinuous changes proposed by saltationists are generally not the result of a great overhaul in the genes. Instead, it is thought that small changes in genotype can result in qualitatively large changes in the phenotype. Thiessen (2009) defines saltational evolution as “a genetic modification that is expressed as a profound phenotypic change across a single generation and results in a potentially independent lineage” (Theissen, 2009, p. 46). Saltations are thus not thought to arise through large genotypic changes, which would be highly improbable, but rather through small changes that involve large effects. Yet as Shapiro (2011) demonstrates, even genotypic variation is not as simple as was once believed: “our view of genome change has become one that describes active cell processes rather than a series of random accidents.” (p. 129) Shapiro calls these processes ‘natural genetic engineering’, which “range from horizontal gene transfers and the movement of transposable elements through chromosome rearrangements to whole genome duplications and cell fusions.” (p. 128) These new discoveries in genetics, coupled with the possibility that small genotypic changes can lead to great phenotypic transformations, show that “clear evidence exists for abrupt events of specific kinds at all levels of genome organisation.” (Shapiro, 2011, p. 128).

The defenders of the saltationist approach all concede that the traditionally upheld small incremental changes exist, but nevertheless attribute a non-negligible role to saltational events. As Theissen (2009) explains: “Advocates of these views often do not completely deny gradual changes (typically during adaptation or microevolution), but consider them unable to explain the origin of phenotypic novelties, or species and higher order taxa.” (p. 44) For some, such as Gould, the role of saltation should not be overstated: “my own betting money goes on a minor and infrequent role” (Gould

2002, p. 1146, quoted in Clark, 2013, p. 182). However as Chouard (2010) has written in a news feature in *Nature*, other saltationists believe that “single-gene changes that confer a large adaptive value do happen: they are not rare, they are not doomed and, when competing with small-effect mutations, they tend to win.” (p. 864) One way or the other, saltations seem to be more and more accepted as a plausible evolutionary scenario, at least in certain cases.

As shall be shown in the next section, Chomsky and his collaborators argue that language was made possible by the appearance of recursion, which allowed the exaptation of the various cognitive mechanisms which comprise the faculty of language (FLB). Insofar as such a minor novelty (recursion) is expected to have yielded such a drastic change in cognitive capacities (language), the approach can be labelled saltationist. The approach is then justified through an argument similar to that advanced by Gould (1980): recursion (defined in section 1.4) cannot be half there; it is either present or it is not, implying that it can only appear in a single step.

1.2. FLN/FLB distinction

1.2.1. The distinction defined

As mentioned earlier, the study of the evolution of the faculty of language is interdisciplinary as language itself calls upon so many aspects of the human organism for its implementation. With so many different disciplines involved, clashes between theoretical frameworks and misunderstandings about the definitions of terms are bound to happen. At the very center of many misunderstandings has been the most fundamental term itself: “language”. No consensus exists regarding its definition: for instance, for some linguists language is defined as the computational mechanisms internal to an individual organism, whereas others argue that language cannot be

divorced from the context in which it is used; for anthropologists or psychologists it is often synonymous with human communication in general. Reacting to this perceived problem, Hauser, Chomsky and Fitch (HCF 2002, Fitch, Hauser and Chomsky 2005; henceforth FHC), propose and refine a distinction ostensibly meant to facilitate cross-disciplinary study of the evolution of language. The faculty of language is thus divided into the two following conceptions:

- (i) FLB, Faculty of Language – Broad, is the group of organism-internal systems used for the production and comprehension of language. FLB also includes within it FLN.
- (ii) FLN, Faculty of Language – Narrow, is the subset of mechanisms in FLB which are uniquely human, and unique to language.

This conceptual distinction is intended to clarify which aspects of language different researchers are focusing on, and defuse any arguments about whether language ought to be defined as communication or as mental computations. As HCF point out: “many acrimonious debates in this field have been launched by a failure to distinguish between these problems” (p.1569). Thus, not only is this distinction made to facilitate communication across disciplines, it allows researchers to continue their respective research programs without treading on each other’s toes. But though this distinction may be ‘politically’ savvy, Churchland (1986) has argued that such an approach may not always be in the best scientific interest. She maintains that in the case of neuroscience and psychology, the disciplines ought not be conceived as tunnelling through a mountain from opposite ends with the objective of meeting one another “because the co-evolution [of the disciplines] typically is far more interactive than that, and involves one theory’s being susceptible to correction and reconceptualization at the behest of the cohort theory.” (Churchland, 1986, p. 373) The same argument applies to any disciplines which work on the same broad issues, as is the case with language evolution. However the distinction is meant to do more than simply resolve differences between disciplines: it is also intended as a guide to empirical research across disciplines.

Indeed, the interest in the delineation of FLN is also to isolate what it is that humans possess which allows us to have language as we do. Since no other animal is able to communicate with human language, it is clear that humans possess certain capacities which are not found in any other species. Although HCF propose contents for FLN (discussed later), the authors do remark in the subsequent article that FLN “could possibly be empty, if empirical findings showed that none of the mechanisms involved are uniquely human or unique to language, and that only the way they are integrated is specific to human language.” (FHC, 2005, p. 181) The conceptual distinction between FLB and FLN thus raises the question as whether there are uniquely human and linguistic capacities and what those capacities are. This in turns leads to theoretical and empirical investigations regarding how and when they might have appeared in our lineage. The interest in FLN is therefore not a hubris-inspired hunch about human exceptionality, but rather an evolutionary-minded question about human-specific adaptations.

1.2.2. FLB and its contents

HCF not only propose the FLN/FLB distinction, but also put forward hypotheses about the contents of each, starting with FLB. As defined above, they propose that FLB includes FLN, but also “at least two other organism-internal systems, which we call “sensory-motor” and “conceptual-intentional”” (p.1570-71). Since both these systems are assumed to *not* be unique to language and humans, HCF point out some promising research in comparative biology in these areas. Regarding the role of the sensory-motor system, HCF applaud research done into vocal imitation and invention in songbirds or dolphins, constraints imposed on vocal tract anatomy in primates, or the capacity for discrimination of sound patterns in language (p. 1573). The conceptual-intentional system for its part involves such aspects as theory of mind and

concept acquisition, with research done on the seeing/knowing distinction in chimpanzees, referential vocal signs in primates, and intentional communication (for a complete list of the research commended by HCF, see Table 1, p.1573). HCF also leave open the possibility that there may be other systems to be included. Although these two systems are not accepted by all as discrete modules within the mind (see e.g. the work in cognitive linguistics: Croft & Cruse, 2004), HCF maintain that these systems which they include in FLB interact in various ways, accounting for human language production and comprehension.

The FLN/FLB distinction may seem fairly straightforward at first glance, but begins to seem a little more confusing when the authors propose that there are parts of the organism which play a supporting role in language production and comprehension, but still don't make it into FLB. According to their conceptualization, FLB "excludes other organism-internal systems that are necessary but not sufficient for language (e.g., memory, respiration, digestion, circulation, etc.)." (HCF 2002, p. 1571) The exclusion of these systems is somewhat intuitive insofar as it is easy to recognize that, for instance, respiration does not play a pivotal role in language production or comprehension (especially when sign language is taken into account). Furthermore, one does want to find a way to exclude parts of the organism that are clearly irrelevant, such as the immune system or a sense of balance. But the characterization of these parts as "necessary but not sufficient for language" is inadequate, since any of the other parts of FLB taken in isolation, including FLN, is necessary but not sufficient for language (since it is taken for granted that it is only through their interaction that we produce language). It furthermore cannot be that all the contents of FLB as understood by HCF are sufficient for language, since clearly memory at the very least is necessary.

If it is confusing to decide what should be excluded from FLB, it is also difficult to settle on the conceptual basis for *inclusion* in FLB. Necessity and sufficiency do not

seem to be the appropriate boundaries for the conceptual distinction if the boundaries are to be drawn where HCF intended them to be. Not only does it give rise to the problems above, it also proves difficult to use when one tries to distinguish which parts of the sensory-motor systems are recruited. For instance, sign language does not involve auditory channels or articulatory muscles, and vocalized language can do without visual cues. What does it mean then to say that the sensory-motor system is part of FLB? What must be intended is some abstraction of the particulars of the sensory-motor system, or perhaps some sort of high-level conversion of the outputs of the other language-faculty modules to whatever motor output is recruited (and presumably the same thing the other way around for the input). Nevertheless, it is perhaps legitimate to speak of the sensory-motor system generally without having to specify the particulars if we assume that there are processes that will be recruited no matter the medium used to communicate. But the question remains: which specific parts of the sensory-motor system should be included within FLB? One can hope that the distinction will serve precisely to clarify such questions, leading to empirical research regarding specific aspects of the systems in FLB which are recruited, in which case a certain amount of vagueness in its delineation is warranted, and assumed to be temporary.

In any case, since the focus of the present work is specifically about the saltationist argument regarding language evolution, which itself relies on FLN and its contents, the problems concerning FLB's unclear delineation need not be resolved here. Fortunately, the FLN distinction is fairly clearly defined, though the same cannot be said of its content, leading to problems which come into sharp focus when seen in the light of language evolution.

1.3. FLN and the saltationist argument

1.3.1. The so-called “recursion-only” hypothesis

The contents of FLN were defined in HCF, but unfortunately in such a way that it has led to much confusion and debate which could have been avoided. In the abstract of the article, HCF “hypothesize that FLN only includes recursion”, understandably leading Pinker and Jackendoff (2005) to characterize HCF’s position as the “recursion-only hypothesis” (p.204). This was an unfortunate coining of the term as it turns out that there never was a “recursion-only hypothesis”. Some years later, Fitch and Hauser both explicitly wrote that they “regret certain editorial decisions that perhaps made our argument more opaque than desirable” (Fitch, 2010, p.75), and put forth a “*mea culpa* for unfortunate cutting of corners in the abstract of Hauser, Chomsky and Fitch (2002)” (Hauser, 2010, p.94). HCF apparently never meant anything like the “recursion-only hypothesis”, and point out (Chomsky, personal communication, Oct. 25th 2013) that they explicitly state in the article that “FLN comprises only the core computational mechanisms of recursion as they appear in narrow syntax *and the mappings to the interfaces*” (HCF 2002, p. 1573, emphasis mine). Recursion thus is meant to interface with the conceptual-intentional and sensory-motor systems in specific ways (and presumably any other system that may be contained in FLB). Those interfaces are thus included in FLN, and are meant to account for certain unique aspects of language.

1.3.2. Critiques by Pinker and Jackendoff

But, of course, what precisely “the mapping to the interfaces” means has also turned out to be problematic. First, Jackendoff and Pinker, in their response to FHC 2005, raise the question as to whether the sentence is to be understood as the “mechanisms

of recursion as they appear in [syntax and the mappings to the interfaces]" or "[mechanisms of recursion as they appear in syntax] and [the mappings to the interfaces]" (Jackendoff & Pinker, 2005, p. 217). In other words, it is not clear whether the sentence is meant to convey that it is recursion which is in both the narrow syntax and the interfaces which is to be considered FLN, or whether the recursion is only in the narrow syntax, and the mappings themselves are not (necessarily) recursive. Second, as Jackendoff and Pinker (2005, p. 217) point out, it is unclear what the mappings to the interfaces are supposed to mean as it is never truly spelled out in either HCF or FHC. As we shall see later, Chomsky does provide more explicit explanations on this front in later articles (but as will also be shown, those explanations raise their own problems).

Jackendoff and Pinker also made it clear that it was possible, as they did, to interpret the characteristics of FLN (*viz.* being unique to language and unique to humans), in a more gradual way than HCF (and FHC). They propose that a given trait may be included in FLN if it has "been modified in the course of human evolution to such a degree that it is different in significant aspects from its evolutionary precursors[...], though not necessarily different in every respect" (Jackendoff & Pinker, 2005, p. 214). Thus, although they adopt the FLB/FLN distinction, they place within FLN many other components of the language faculty, such as phonology, certain parts of concept formation, words, speech perception, etc. With respect to phonology for instance, Pinker and Jackendoff state: "It appears that some of the combinatorial properties of phonology have analogues in some species of birdsong, and perhaps in some cetacean song, but not in any primates; if so, they would have to have evolved separately in humans" (Pinker & Jackendoff, 2005, p. 212). FHC for their part reject this approach in favour of a more strict delineation, all the while conceding that other researchers have defined FLN differently, "leaving the possibility of a more inclusive definition open to further empirical research" (HCF 2002, p.1571).

1.3.3. Language's appearance through saltation

Such restrictions on admittance to FLN allow HCF to advance claims that it is not the result of adaptation but rather of a single random mutation. Whereas the contents of FLB may have “an ancient evolutionary history long predating the emergence of language”, the hypothesis that FLN contains only recursion allows them to argue that “if FLN is indeed this restricted, this hypothesis has the interesting effect of nullifying the argument from design, and thus rendering the status of FLN as an adaptation open to question” (HCF 2002, p.1573). In other words, all prerequisites for language were already in place when recursion appeared, thus allowing the components of FLB to be exapted for language, implying that the faculty of language was not shaped as an adaptation for language. HCF leave the door open as to the origin of the sudden use of recursion in language, stating both that it “appears to lack any analog in animal communication and possibly other domains as well” (p. 1571)—suggesting that it was a novelty in the biological world at the time of its appearance—and that it may be used “to solve other computational problems such as navigation, number quantification, or social relationships” (p. 1578) in humans or other species—suggesting that it may be an exaptation itself. The latter seems more plausible insofar as the capacities of navigation and social relationships certainly are present in other species, but it needs to be shown that they too rely on recursion.

The saltationist argument also implies that language could not have evolved gradually, becoming more and more complex as adaptation ran its course, but rather appeared at a single moment in our species' history. This also precludes the possibility of a proto-language, on the premise that language is defined as the whole of FLB, including FLN. Thus, even if pre-linguistic communication could be shown to exist in hominids, the definition of language advanced by HCF is such that without

FLN, any talk of mechanisms related to communication cannot be referring to language.

HCF therefore implicitly hold a saltationist approach to language evolution, as opposed to a gradualist one (as we shall see in chapter III, in subsequent publications Chomsky has made his saltationist approach more explicit). As explained earlier, if no intermediate steps can be imagined between a trait's absence and its apparition, then there is a good chance that it is the result of saltation. HCF claim that this is the case when stating that "we see little reason to believe either that FLN can be anatomized into many independent but interacting traits, each with its own independent evolutionary history, or that each of these traits could have been strongly shaped by natural selection" (p.1574) In contrast, if FLN contains only a single component, it is plausible to maintain that it appeared through one random mutation, especially if that component is thought to be very simple, yet able to account for the sudden appearance of a fully-formed language capacity.

HCF unfortunately never define recursion, despite the fact that it plays such a central role in their theory, and according to them, in language itself. The closest they come to defining it is by explaining that recursion "takes a finite set of elements and yields a potentially infinite array of discrete expressions" (HCF, p. 1571), without explaining how this is done. In the next section I will explore various definitions and show that the choice of definition has a strong impact on the claims in HCF.

1.4. Recursion defined

The notion of recursion is used in many disciplines, and its precise definition often differs depending on the context. Generally speaking, recursion is a process whereby an object, image, phenomenon, etc. calls upon or creates itself anew. A facile

illustration of this is to be found in an image which contains within itself the very same image, which itself contains the very same image, and so on. In computer science, a function or algorithm is recursive when the process calls upon itself in its own generation (Fitch, 2010). In linguistics, recursion has often been assumed to be central to language production, yet until recently, rarely rigorously and formally defined. As has been pointed out many times (cf. Parker 2006, Tallerman et al. 2009, Karlsson 2010, Fitch 2010), even linguistic recursion can be defined in many ways. The following are two distinct definitions which are both employed in linguistics.

1.4.1. Recursion – definition 1: embedding

A definition that was long used (often implicitly) in generative grammar is the following, sometimes referred to as “self-same embedding” and hereafter named “recursion-embedding”:

- (iii) **Recursion-embedding:** the capacity for an algorithm to insert a structural unit of a particular type (e.g. noun phrase, sentence) within another unit of the same type (Tallerman, *et al.*, 2009, p. 140).

This is the type of recursion which was found in Chomsky’s earlier writings, at least as early as *Syntactic Structures* (1957), where the explicit goal was “the construction of a grammar that can be viewed as a device of some sort for producing the sentences” (p. 11) of a given language. That “device” takes the form of a set of rewrite rules which constructs sentences from the top down. The following illustration is proposed by Bickerton (2009a, pp. 241-242): a given sentence S can be subdivided into a noun phrase (NP) and a verb phrase (VP). Thus we have the rule:

- (iv) $S \rightarrow NP VP$

These phrases as well contain various elements, such as determiners (Det), prepositional phrases (PP), as well as verbs (V) and nouns (N), yielding the following rules (items in brackets are optional):

- (v) $NP \rightarrow (Det) N (PP) (S)$
- (vi) $VP \rightarrow V (NP (NP) (PP) (S))$
- (vii) $PP \rightarrow P NP$

These are recursive rules since S contains within it NP and VP, which each in turn contain S, implying that sentences can be inserted in sentences. A NP can also be contained within a NP through the PP rule. This type of recursion is found in sentences such as (the notation is simplified for ease of reading):

- (viii) [John thinks that [Mary loves Jake.]]_S
- (ix) [The dog on [the dock]_{NP}]_{NP} is noisy.

This set of rules therefore allows the production of an infinite set of sentences through finite means, since the recursion ensures that rules can be repeated infinitely.

1.4.2. Recursion – definition 2: cyclical

A second way to define recursion has become popular in the Minimalist Program in linguistics (cf. Nevins, Pesetsky, & Rodrigues 2009, Rizzi 2009). Researchers in the program posit that one of the central, if not the only, mechanism in language is ‘Merge’, which joins two units into a single unit, and can potentially do so an infinite number of times, using its own output as input (Bickerton, 2009c). According to this approach, recursion has the following broader definition, hereafter referred to as “recursion-cyclical”:

- (x) **Recursion-cyclical:** “the property whereby certain formal rules or rule systems reapply to their own output” (Tallerman *et al.*, 2009, p. 140)

Merge can join previously Merged units with other units, meaning that it can be applied to its own output indefinitely, and as such, is a recursive process. Note that Merge is not the only process that abides by recursion-cyclical: by definition, any process which can take its previous output as input is considered recursive. Unlike recursion-embedding, this type of recursion creates a bottom-up approach to sentence

construction since it begins with the basic lexical units and constructs sentences by adding additional units. Merge and this type of recursion will be discussed in detail in section 2, as it has come to be central to Chomsky's views on language evolution.

1.4.3. Effects of recursion

Whether relying on definition 1 or 2, an important effect of recursion is that it is able to produce a boundless number of sentences using finite means. This point has been raised countless times by Chomsky (eg. 1957, 1966, 2000a, 2005), as well as in HCF (p. 1571) and is seen as an essential feature of the faculty of language. Of course within anyone's finite lifetime it is impossible to produce or observe an infinity of sentences, and it must be inferred from available data and starting assumptions. By definition, both definitions of recursion are combinatorial processes which join discrete units (words or parts of sentences) into sentences which are boundless.

However it needs to be noted that other types of algorithms, in particular iteration, can also create a boundless number of strings from finite means. An iterative algorithm is one which has the capacity to repeat its own process potentially infinitely, just as a recursive one does. Indeed the distinction between these two types of algorithms can be difficult to grasp, since one might be tempted to think of one as a subclass of the other. However that is not the case, as each proceeds in different ways, yielding significant differences.

The most significant difference is recursion's capacity for building a hierarchical structure within its constituents. Indeed it has long been recognized that sentences adhere to a hierarchical structure, thus accounting for the long-range dependencies between parts of sentences (e.g. subject-verb agreement, reflexives, etc.). A recursive algorithm builds this type of structure by including units within units, creating a

hierarchy. As Karlsson (2010) explains, an iterative algorithm, for its part, will not produce a hierarchical structure, yielding instead “flat output structures, repetitive sequences on the same depth level as the first instance.” (p. 2) It follows, as Rizzi (2009) points out, that mere iteration of a process is not sufficient to claim it is recursive despite the fact that it too could produce a boundless number of sentences using finite means. According to him, it is also “its capacity to create a hierarchical structure” (p.65) which is tell-tale of the recursive nature of a given algorithm. One important thing to remember however is that any structure created in language is only ever inferred, since externalized language is always serialized and therefore transformed into a flat structure for output. This compounds the difficulty in differentiating between a recursive and an iterative algorithm since, as shall be explained in the next section, in most cases the serialized output of an iterative algorithm can be made to match that of a recursive one.

1.5. Critiques to recursion-only

In this section I will explain two critiques that have been levelled at the recursion-only hypothesis, and show how they are not as potent as the authors expected them to be. The first is some corpus analysis done by Karlsson (2010), wherein he shows that the outputs of recursion-embedding are for the most part able to be reproduced through mere iteration. Furthermore, those examples which are unambiguously recursive-embedding are exceedingly rare, implying that recursion cannot be a central part of language. The second critique is Dan Everett’s well-known and controversial claim that there exists a language, Pirahã, which does not contain recursion. As we shall see, both these attempts to rebut the recursion-only hypothesis fail to do so, among other reasons because they attack a straw man in assuming that recursion can only be defined through embedding, as described in definition 1 above. As will be shown in chapters II and III, Chomsky explicitly adopts recursion-cyclical in his later

publications (Chomsky, 2010; Berwick & Chomsky 2011), rendering critiques to recursion-embedding obsolete (whether the choice of recursion-cyclical is superior is another matter: see Everett, 2009, p. 437-438 for a brief discussion regarding this issue). Nevertheless, at the time these critiques of HCF seemed important, still raise important points, and so are addressed here.

1.5.1. Karlsson: "True" recursion is rare

Karlsson (2010) claims that the recursion-only hypothesis does not stand up to scrutiny because it simply is not manifest enough in spoken and written language. The first thing Karlsson notes is that much of recursion can be mimicked by iteration. Indeed, algorithms that specify sequential arrangement can come in at least two forms: recursive and iterative. As mentioned earlier, whereas a recursive process will call upon its own output to yield a hierarchically structured output, "iteration yields flat output structures, repetitive sequences on the same depth level as the first instance." (Karlsson, 2010, p. 2) Yet the difference between the two is not obviously apparent in the output, since externalized language is always linear, and the depth of structure (should there be any) can only be inferred through analysis. In light of this, Karlsson claims that only very specific kinds of recursion-embedding are not convertible to iteration. Recursion-embedding comes in two subtypes: nested recursion and tail-recursion. Tail-recursion adds components to the left or the right of a given structure; nested recursion 'splits open' the existing structure to add components in the middle. Sentence (viii) above is an example of tail recursion, repeated here for convenience:

(xi) [John thinks that [Mary loves Jake.]_s]_s

The following is an example of nested recursion:

- (xii) [A lot of the housing [that the people [that worked in New Haven] lived in] was back that way.] (Reich and Dell 1976, quoted in Karlsson 2010, p.8)

According to Karlsson, only nested recursion is not algorithmically convertible to iteration, and therefore “nested recursion (center-embedding) is recursion par excellence” (Karlsson, 2010, p. 6).

Karlsson then goes on to show through analyses of large corpus data of European languages that nested recursion is in fact quite rare, whether in written or spoken language. The maximum level of center-embedding for written language is 3, and for spoken 2, and appears so rarely in spoken language (twice in the hundreds of corpuses analysed) that Karlsson considers it “practically absent from ordinary spoken language” (p. 8). He then concludes that recursion “cannot therefore reasonably be considered a central design feature of language, as claimed by Hauser, Chomsky, and Fitch (2002)” (Karlsson, 2010, p. 8) since in the great majority of cases it is replaceable by iteration.

1.5.2. Critique of Karlsson

Two possible rebuttals from Chomsky come to mind, which Karlsson effectively dodges. The first of those criticisms regards the relevance of corpus analysis to linguistics. Chomsky’s criticism in this respect relies on the well-known competence/performance distinction he introduced early in his career. The distinction casts serious doubt on the relevance of research done on corpuses, since analyses of performance cannot be directly transferable to competence. However in the context of evolution this distinction does not apply directly: natural selection cares not a whit about competence, and selects only on the basis of performance. Research into performance is thus very relevant to language evolution. The second critique relates to Chomsky’s emphasis on parsimony in explanations in linguistics (a desideratum

especially apparent within the Minimalist Program, as is seen in chapter II): since there do exist cases of center-embedding which all typical speakers are capable of producing and understanding, and since recursion captures not only those cases which iteration can cover but also the cases of center-embedding (no matter how sparse they may be), it is more parsimonious to assume that the algorithm is only recursive, rather than assume that it is at times recursive, at times iterative. Once again, evolution does not care for the details regarding the mechanisms underlying performance, and does not in principle favour parsimony of competence.⁴

Karlsson's critique rests on a very different paradigm for linguistics, which assumes that the structural principles proposed by Chomsky and others within the generative enterprise can all be translated to principles calling only on linear sequence. Since iteration yields only flat output structures, accounting for all syntactic phenomena necessarily uses principles of basic linearization, free modification, and information-packaging (Karlsson, 2010, pp. 1-2). This implies that the claim that iteration can replace recursion in practically all cases rests on very different premises. The relative value of these paradigms is hotly debated by proponents of each, with no clear winner. The significance of Karlsson's critique thus rests on validation of his more general linguistic theory, which remains open to debate.

Finally, it is interesting to note that the entire premise of the criticism rests upon the idea that linguistic recursion is only recursion-embedding. As will be shown in chapter II, Chomsky, at the very least, has espoused a view of language evolution that rests entirely on recursion-cyclical, meaning that Karlsson's approach misses the mark if Chomsky's theory was the aim.

⁴ Of course, as Parker & Maynard Smith (1990) have argued, evolution does tend toward optimality, but it is optimality of performance that is in question, not competence.

1.5.3. Everett: Pirahã

In 2005, Dan Everett published an article in *Current Anthropology* wherein he claimed to have discovered a language with no recursion. The language is spoken by the Pirahã, a small tribe in the Brazilian Amazonian jungle, which Everett visited for over 30 years, even living there with his family for years. His findings have led to a long and often acrimonious debate surrounding the validity of his claims, and its implications regarding language evolution.

The Pirahã language is a very complicated issue. When Everett's (2005) article first came out, it seemed that he had perhaps found a drop-dead argument against HCF 2002. The argument went as follows: if a language exists without recursion, then recursion cannot be the sole component of FLN, since if FLN has only a single component, it must necessarily be present in every language. Otherwise, since by definition FLN contains that which makes human language unique, there would be no difference between a non-recursive language and a non-human way of communicating. The claim the Everett had found such a language therefore caused a stir, and drew fire from some linguists.

One salient critique that has been levelled at Everett is that his analysis of Pirahã language is faulty. Nevins, Pesetsky and Rodrigues (2009) argue that the analysis done in Everett (2005) contradicts earlier analyses by Everett himself which show the presence of recursion-embedding in Pirahã. They furthermore claim that most of the oddities pointed out by Everett regarding the language are "illusory, nonexistent, or not supported by adequate evidence" (Nevins, Pesetsky, & Rodrigues, 2009, p. 356).

Needless to say, Everett replied (2007⁵, 2009), but since he is the only trained linguist to master the language, the field of linguistics is at pains to validate or deny his claims.

1.5.4. The “official” answer and its pitfalls

The “official” answer, found in HCF’s follow-up article, FHC 2005, was that the existence of languages with no recursion is “irrelevant to the questions under discussion”, since “the putative absence of obvious recursion in one of these languages is no more relevant to the human ability to master recursion than the existence of three-vowel languages calls into doubt the human ability to master a five- or ten-vowel language” (Fitch, Hauser, & Chomsky, 2005, p. 203). Evoking Jackendoff’s (2002) analogy of language calling on a toolkit of cognitive resources, the authors argue that recursion is one tool among many that enables language acquisition, and that “not all languages use all the tools” (Fitch, Hauser, & Chomsky, 2005, p. 204; see also Bickerton, 2009a, p. 239 for a similar response to Everett).

This answer is not satisfactory as-is, beginning with the toolkit analogy. Insofar as FLN contains only a single item, that item cannot be one optional tool among many, to be used or not depending on the language. As argued above, if recursion is the only component differentiating human language from other forms of communication, its absence in a particular human language implies that its inclusion in FLN is problematic. At the very least, it entails either that there must be more than one component in FLN (thus rehabilitating the toolkit analogy), or simply that recursion is not a component unique to language, in other words is part of FLB, not FLN. It

⁵ This predates the Nevins *et al.* (2009) publication as it was a response to the circulated, not yet published version.

may be necessary to point out here that this critique still holds if one remembers that the “mappings to the interfaces” need be included in FLN. The mappings themselves cannot provide the additional tools for the toolkit, since recursion is the sole component in FLN which can determine those mappings. This therefore implies that recursion must still be present and operational for the mappings to be formed. Thus, if the toolkit analogy is to be employed, FLN must either contain more than one item, or the analogy applies to the various items in FLB.

The second, implied answer is slightly more convincing, but not without its own problems. Implied is the fact that the absence of “obvious recursion” is not proof of absence of recursion in the cognitive process of language. Indeed, as mentioned before, once language is externalized, all output is serialized and therefore all trace of recursion is only inferred. Thus, in order to show that a given language truly is without recursion, one would need to show that recursion is absent at all steps of the generative process. By Everett’s own admission, the Pirahã people are able to think recursively: “I claim that the Pirahãs lack recursion in the syntax. I made no such claim about their semantics or their discourse, for example. In fact, I have given many examples of recursion in discourse as different ideas are contained in others within the main and subordinate story lines.” (Everett, 2007) Thus the important distinction is between linguistic capacities and general cognitive capacities. If recursion can be shown to simply be a part of general cognition and not tied specifically to language, then it cannot be in FLN, since by definition FLN contains only components unique to language. However to advance such a claim, one would need to maintain simultaneously that (1) speakers of Pirahã think recursively only within the faculty of language and not general cognition (by definition of FLN), and (2) that their inability to express that recursivity through language tells us nothing about recursion in their language faculty. Such a position is difficult to imagine, and even more difficult to demonstrate empirically.

So it would seem that, should Everett's analysis of Pirahã be accurate, he does indeed have quite a strong bit of evidence against the recursion-only hypothesis. But, as should be clear from the discussion on recursion in the previous section, the weight of Everett's claim depends on the definition of recursion he adopts. Everett claims more specifically that Pirahã lacks recursion-embedding. Pirahã are therefore purportedly unable to construct sentences such as "Mary thinks that John is nice", where a sentence is embedded in a sentence. Although this may be an interesting finding in and of itself, it has become apparent that of the three authors, Chomsky, at the very least, did not mean for recursion-embedding to be the kind of recursion found in FLN as his papers since then all focus on Merge and recursion-cyclical. As Nevins et al. (2009) point out, if recursion-cyclical is adopted, "a language that lacks recursion would be considerably more exotic. No sentence in such a language could contain more than two words" (footnote 11, p.366). Needless to say, Pirahã does produce sentences with more than two words, making it a facile argument that they do indeed use recursion-cyclical. Of course this may mean that recursion-cyclical is a type of recursion that is too broad to be restricted to FLN, an issue tackled in chapter II, section 2.6.

But before developing Chomsky's shift towards recursion-cyclical, let me first propose my own critique of the recursion-only hypothesis, which I believe to be superior to the two presented here since it neither relies on corpus data and all the problems that can bring about, nor on a single linguist's analysis of a language, nor even on a specific definition of recursion.

1.6. Recursion as a property

The most significant problem with the hypothesis advanced in HCF 2002 is that it is underspecified insofar as it does not put forward any specific mechanisms to account

for language competence. This becomes clear with a closer look at the role of recursion in the so-called “recursion-only hypothesis”, and the possible interpretations of their claims.

The first thing to note is that “recursion” is not, properly said, an object in and of itself (“object” here is meant to be theory-neutral and denote whichever term is deemed appropriate, be it algorithm, module or whatever type of process is expected to fulfill that role in language competence). It is rather a property which an object may instantiate. As was described earlier, recursion, defined generally, can be instantiated by many things, including illustrations, formulas, algorithms or any generative process. The specific definitions found in linguistics as well can be instantiated by various algorithms, and no algorithm can have recursivity as its sole property since it must (at least) also specify input and output conditions. In other words, there is no such thing as a ‘purely’ recursive operation, or *the* recursive operation

1.6.1. Recursion instantiated

Considering the fact that recursion must be instantiated in an object, one can interpret the “recursion-only hypothesis” in one of two ways. Either (a) it claims that only the recursive aspect of a given algorithm is unique to language and to humans, or (b) it is simply a condensed way of saying that FLN contains only an algorithm which must be recursive. Neither of these approaches works, for different reasons.

In order to defend (a), one would need to show that all properties of the algorithm, with the exception of recursion, are not unique to language and to humans (by definition of FLN). This would imply explaining how the very “same” algorithm, omitting the recursive aspect (in some strange way of interpreting “same”) is part of

FLB. This seems conceptually impossible, as on the one hand a recursive algorithm without recursion can hardly be conceived of as the same algorithm since it would specify very different input-output combinations. On the other hand, it is difficult to conceive of what is left in FLN since the recursive property alone does not specify inputs and outputs of FLN, which are needed if we are to understand FLN as part of the generative process of the faculty of language.

Certain passages in HCF hint that interpretation (b) seems to be closer to what HCF intended, but the object/property distinction is never truly addressed. The following passage suggests—albeit not entirely clearly—that they understood FLN as containing computational mechanisms which must be recursive (underlined sections will be addressed following the passage):

Without prejudging the issues, we will, for concreteness, adopt a particular conception of this architecture [of the faculty of language]. We assume, putting aside the precise mechanisms, that a key component of FLN is a computational system (narrow syntax) that generates internal representations and maps them into the sensory-motor interface by the phonological system, and into the conceptual-intentional interface by the (formal) semantic system; adopting alternatives that have been proposed would not materially modify the ensuing discussion. All approaches agree that a core property of FLN is recursion, attributed to narrow syntax in the conception just outlined. (HCF 2002, p. 1573, emphasis mine)

This passage is very difficult to parse because of its dissonance with other claims made elsewhere in the article. First, the proposed contents of FLN are now referred to as “key components of FLN,” leaving open the idea that FLN may contain more than simply recursion or a recursive mechanism. Second, that key component is not recursion but an unspecified syntactic, generative process of internal representations. Third, recursion is here explicitly referred to as a “core property” not of the computational mechanism, but of FLN itself, implying that FLN itself is the object instantiating the property of recursion, rather than being a conceptual distinction between parts of the language faculty. To be fair, in this case the authors may have

meant FLN to refer to the mechanisms themselves rather than the distinction. These three points lead to believe that the “recursion-only hypothesis” meant for FLN to contain not recursion alone, but a mechanism which must be recursive. But this does not circumvent the most significant problem: what mechanisms are we talking about specifically?

1.6.2. The recursive property and the interfaces

HCF apparently assume that they can put aside the question of the precise mechanism since “all approaches agree that a core property of FLN is recursion”, but it nevertheless leaves their hypothesis peculiarly unspecific. Needless to say, determining which recursive mechanisms instantiate recursion will inform much of the discussions surrounding linguistic competence, the cognitive architecture involved in language, the “interfaces to the systems”, and even language evolution itself. Taking for instance the “interface to the systems” which HCF include in FLN, it is clear that those interactions between the sensory-motor system and the syntactic system will be determined in no small measure by the specific inputs and outputs of the algorithm which accounts for syntactic competence. In other words, since recursion is a property and not an algorithm or mechanism in and of itself, it is incoherent to assume that only that property and its interfaces would comprise FLN, since whatever algorithm or mechanism that is instantiating the recursion must also be in FLN, and it is that algorithm which would determine the interfaces to other systems, rather than merely its recursive property.

It is possible that HCF deliberately proposed a weak hypothesis in order to be inclusive of the most linguistic theories possible. FHC, their second article, does seem to support this interpretation: “To be precise, we suggest that a significant piece of the linguistic machinery entails recursive operations” (p.182); and later: “The only

assumption made in HCF, and here, about syntactic theory is the uncontroversial one that, minimally, it should have a place for recursion.” (p.183) But if the “recursion-only hypothesis” in fact meant that FLN must contain algorithms which minimally have recursion as a property, then the hypothesis is so weak as to do little to advance research since HCF themselves point out that “all approaches agree” with this claim already (though not everyone agrees with this claim). By diluting the hypothesis to this extent, it leaves it open as to which mechanism plays the key role of FLN since linguists have already proposed various recursive mechanisms to account for language production and comprehension. Consequently, the hypothesis in HCF does little to restrict research into language evolution any further than it already was, except to assume that some as-yet-unspecified syntactic component is unique to humans and to language. This, incidentally, is just what Chomsky’s hypothesis of Universal Grammar has always been. If the principal aim of the hypothesis was to aid in circumscribing the precise mechanisms which should be at the heart of investigations into language evolution, then the article has missed its mark.

1.7. Conclusion

In sum, HCF’s article featured Chomsky’s explicit entrance into the language evolution debate, but suffered from many problems which could have been avoided with careful attention to the concepts deployed. FLN and FLB could potentially serve as useful tools for communication across disciplines, but would need to be more rigorously defined to avoid inclusion or exclusion of unwarranted mechanisms. The content of FLN is the most contentious and problematic aspect of HCF, with claims regarding the status of recursion that at times seem contradictory, as well as a lacking but necessary definition of recursion. This left them open to many critiques which unfortunately ended up attacking a straw man once recursion and the “recursion-only hypothesis” were better clarified. Finally, once it is recognized that recursion is a

property, it becomes clear that recursion cannot be the sole component of FLN since it must be instantiated. At best, HCF merely meant that FLN must contain a recursive algorithm, in which case it is merely a rehashing of Chomsky's argument for Universal Grammar.

The next chapter explains Chomsky's current favoured approach to linguistics, the Minimalist Program, including the mechanisms advanced to account for the language faculty. These specific mechanisms, as well as the notions advanced in HCF regarding the FLN/FLB distinction and the emphasis on recursion, lay the foundations for Chomsky's more recent approach to language evolution as explained in chapter III.

CHAPTER II

THE MINIMALIST PROGRAM

2.1. Beyond the “recursion-only” hypothesis

Following the publications co-authored with Hauser and Fitch (Hauser *et. al.*, 2002; Fitch *et. al.* 2005), Chomsky published articles which advanced slightly different proposals regarding the evolution of language, most explicitly in Chomsky (2010) and Berwick & Chomsky (2011). These articles take as foundation the work done within the Minimalist Program (henceforth MP), the latest paradigm adopted by Chomsky for linguistic investigations. Since the MP advances specific mechanisms to account for the workings of the language faculty, the articles concerning language evolution do so as well, answering the main criticism levelled at HCF (2002) in chapter I. HCF (2002) proposed that FLN be comprised of only recursion and the interfaces to the CI and SM systems, a proposition I argued was incomplete once it is recognized that recursion is a property which must be instantiated in an algorithm, implying that the algorithm in question as well must be included in FLN. By adopting the specific algorithms and properties put forward by the MP, Chomsky (2010) and Berwick & Chomsky (2011) resolve the problem, but as shall be seen in chapter III, open themselves to additional important critiques.

In order to understand the proposals advanced in Chomsky’s later writings regarding language evolution, a cursory knowledge of the MP is necessary. This chapter is therefore a general exposition of the main mechanisms proposed within the MP to account for the workings of the language faculty. I begin with a very brief mention of

the history and aims of the MP in section 2.2, followed in section 2.3 by a description of the basic mechanisms proposed within the MP framework, namely Merge and the lexical items. Section 2.4 puts the mechanisms to work, showing how a syntactic derivation is performed, and the complexities that arise. In section 2.5, I investigate the place of Merge within the faculty of language, arguing that Merge's simplicity depends on offloading computational complexity—and even recursion itself to some extent—onto lexical features. Finally, section 2.6 investigates the possible broader application of Merge in human cognition and non-human capacities.

2.2. The Minimalist Program as field of research

2.2.1. Brief history

Chomsky has often changed his linguistic theories over the years, building upon new ideas and often leaving behind former students or colleagues beleaguered in a paradigm he now considers obsolete. His most recent such shift has been with the development of the MP, first proposed in an article in 1993, re-published along with other articles in Chomsky's book *The Minimalist Program* (1995). In a nutshell, the MP tries to reduce the linguistic mechanisms proposed, and thus the complexity of the language faculty, to a minimum.

The MP is said to have grown out of the “Principles and Parameters” (P&P) approach, which posited that language acquisition is regulated by broad principles which are ‘activated’ or not by certain parameters. Chomsky explains the theory through a switch metaphor:

We may think of the language faculty as a complex and intricate network of some sort associated with a switch box consisting of an array of switches that can be in one of two positions. [...]. The fixed network is

the system of principles of universal grammar; the switches are the parameters to be fixed by experience. (Chomsky, 1988, pp. 62-63)

Exposure to the language of the surrounding community makes the individual's language faculty switch the various parameters, permitting acquisition of any and all languages of the world. Berwick (2011) explains that "akin to atomic theory, this small set of constraints may be recombined in different ways to yield the distinctive syntactic properties of diverse natural languages, just as a handful of elements recombine to yield the many different molecular types" (p. 82) The P&P approach assumed that the syntax-building algorithms were many, and interacted in various ways to yield all the possible human languages. In this view the faculty of language remained a complex phenomenon with a heavy computational load on the syntactic algorithms, a notion the MP set out to disprove.

The MP grew out of the P&P approach, but the reason why is debated. For Holmberg (2000) the MP stems from certain empirical problems encountered in the P&P approach, licensing the paradigm shift. For others (Boeckx, 2006) the MP is the result of the successes of the P&P, with a new emphasis on optimality: "solving the acquisition problem was the paramount measure of theoretical success in linguistics [for the P&P approach]. Once, however, this problem is taken as essentially understood, then the question is *not* how to solve it but how to solve it *best*." (Boeckx, 2006, p. 61) For others still (Lappin, Levine, & Johnson, 2000), the MP rests not on empirical findings and difficulties stemming from the P&P approach, but rather on obscure appeals to efficiency and optimality, driven mostly by Chomsky's speculations and his strong influence in the field. Whatever its origins, the MP represents a shift away from the complexity found in the P&P approach to a more parsimonious view of the faculty of language.

2.2.2. Aims

The MP's general purpose is said to be the reduction of the number of mechanisms needed to account for language. According to Chomsky (2010) the MP "is largely theory neutral" and applicable "whatever one's beliefs about design of language may be" (p. 51). Of course this statement is debatable, as the MP is far from unanimously adopted in linguistics, as demonstrated by the many publications criticizing the program or proposing alternative paradigms. As is shown in section 2.3, the MP does not limit itself to broad ideals of parsimony, but also proposes more or less concrete and precise mechanisms to account for the language faculty.

The MP is deliberately presented as a program and not a theory to emphasize the fact that it is meant as a general line of inquiry to be pursued, and is furthermore ostensibly theory-neutral. Evoking Imre Lakatos' notion of "research programme" in philosophy of science (Boeckx, 2006, pp. 6-7), the MP proposes a core idea which remains even if the more superficial auxiliary hypotheses could eventually be disproved. Di Scullio and Boeckx (2011) characterize that core as a "challenge to the linguistic community: Can it be shown that the computational system at the core of the language faculty is optimally or perfectly designed to meet the demands on the systems of the mind/brain it interacts with?" (pp. 3-4) The systems of the mind in question are those found in HCF (2002): the Conceptual-Intentional (CI) and Sensorimotor (SM) systems. The components of the language faculty thus must link in the most direct way possible the two systems, "containing a minimum of specific grammatical machinery" (Benítez-Burraco & Longa, 2010, p. 315). Boeckx (2006) describes the idea of a 'perfect' design for the language faculty as the strongest hypothesis, "probably too strong", which "acts as a limiting case, enabling us to see more precisely where and when the hypothesis fails and how much of it may be true." (Boeckx, 2006, p. 4) Researchers thus attempt to resolve the problems raised by

linguistic phenomena using as few stipulations and mechanisms possible, in the hopes of demonstrating that the assumption of a “perfect” language faculty is warranted.

The stated purpose of the emphasis on efficiency and parsimony within the MP is a belief that it will lead to reconciliation with the biological sciences. This approach, alternately dubbed ‘biolinguistics’, “forces linguists to reformulate previous findings in terms of elementary units, operations, and interface conditions” (Di Scullio & Boeckx, 2011, p. 4), concepts which it is believed will be found in other biological domains. These parallels with other fields could then be used to “contribute to our understanding of how core properties of language are implemented in neural tissues and how it [*sic*] evolved in the species.” (Boeckx, 2011, p. 43) By reducing mechanisms to a minimum and appealing to broad computational principles, researchers in the MP hope to end “linguistic isolationism” (Di Scullio & Boeckx, 2011, p. 4) and show that linguistic mechanisms are amenable to biological and neurological investigation through discoveries of parallels or correlates.

This explicit aim raises some questions regarding the plausibility of such a research program insofar as the mechanisms that stem from biological evolution need not—and perhaps even tend not—to be parsimonious in their operation. Although adaptations can tend towards optimality, that optimality is geared towards fitness (Parker & Maynard Smith, 1990), and not the inner mechanisms; because the mechanisms are more often than not the result of the tinkering of evolution, they may end up being less than elegant, yet nevertheless appropriately functional for the purposes of fitness. Thus the biolinguistics approach as it is stated above is potentially problematic and may not mesh properly with evolutionary biology (see Kinsella & Marcus 2009 for more detailed arguments, and Narita & Fujita 2010 for a response). These questions are all the more relevant since, as will be shown in chapter III, the search for minimalism plays a central role in the saltationist justification. However the aim of this paper is to critique the saltationist position, taking for

granted the theoretical framework built by Chomsky and his collaborators, and as such, these interesting questions will be set aside for the time being.

2.3. Building blocks of the MP

Notwithstanding the emphasis on the fact that the MP is supposed to be a program and not a theory, linguists within the program adopt a set of assumptions regarding basic elements of the language faculty. Theories proposed may vary with respect to details, but all seem to agree that there exist at least two elements needed to account for the language faculty: Merge and lexical units (Clark, 2013, p. 188). Both are briefly described below in sections 2.3.1 and 2.3.2, with their full role and interaction shown in section 2.3.3.

2.3.1. Merge

Merge is the basic combinatorial algorithm within the MP, and is expected to account for all syntactic manipulations. It is put forward as “the basic structure building operation of syntax” (Collins & Stabler, 2011, p. 5) or “the basic syntax-creating process” (Bickerton 2009b, p. 4). Merge combines two lexical units to create a single lexical unit containing the two original units. It can be described as “an operation that takes structures X and Y already formed and combines them into a new structure Z [...] we can take Z to be simply {X,Y}” (Chomsky, 2010, p. 52). The units thus “Merged” (when capitalized, I refer specifically to the action performed by Merge, or result thereof) remain intact, and the new lexical unit is merely the set of the two original units.

Merge can be further subdivided into (at least) two distinct types of Merge: *External* and *Internal* Merge. External Merge is the straightforward type described above, where two lexical units are combined into a set. For instance, in the case where Merge applies to the (previously Merged) set {A, B} and the unit C, the result of such an External Merge is {C, {A, B}}. A more concrete example is the External Merge of {eat, that} with {you}, yielding the phrase {you, {eat, that}}. Internal Merge is used in a case where one of the units to be Merged is a subset of the other item to be Merged (Chomsky, 2010, p. 54). For example, if Merge is applied to the (previously Merged) set {A, B}, and one of the units within it, say B, the necessary operation is Internal Merge, with the resulting set: {B, {A, B}}. This latter type of Merge is often used in question formation, as in the case of Merging {will} with {you, will}, yielding {will, {you, will}}, where “will” is not only a second application of the word “will”, but actually the same token of the word. At a later stage, before being externalized, the first “will” is suppressed, yielding the question “will you?”⁶ The difference between the two types of Merge is the origin of the items to be Merged: for External Merge, both items are independent of one another (either drawn from the lexicon or taken from previous applications of Merge), whereas for Internal Merge one item is already contained within the other. The use of both types of Merge is explained in greater detail in section 2.4, with additional examples. According to Rizzi (2009), the two types of Merge should not be seen as distinct operations: “What differs in the two cases is simply the way in which the two candidates of Merge [B or C], are selected through search in the available computational space” (p.80). Thus

⁶ Some explanation may be needed for those less familiar with generative grammar: “will” is assumed to follow “you” since it is interpreted as the verb of the subject “you”, but it is also thought to also precede “you” since that is how it is expressed once externalized. It was previously thought that the word “will” moved to the beginning of the sentence at some point in the derivation, but it is now thought to be present at both positions, and only later suppressed at the original position. Internal Merge is thus meant to account for phenomena previously accounted for through “move” in earlier theories of generative grammar.

Merge, despite these two ways of applying it, is nevertheless thought to be a single operation.

As described in chapter I, Merge is therefore recursive-cyclical: it can apply to its own output. Indeed it can call upon any lexical unit to combine with another, but also those lexical units which were created by previous applications of Merge. There is hence no limit to iterations of Merge, yielding, as sought, a language which is “a system of discrete infinity” (Chomsky, 2005, p. 11). As is desirable within the MP, Merge is expected to be “the simplest mode of recursive generation” (Berwick & Chomsky, 2011, p. 30), or “the ultimate distilled format of syntactic recursion” (Rizzi, 2009, p. 67), satisfying the requirement for a recursive mechanism, all the while remaining as elementary as possible.

2.3.2. Lexical items

For Merge to be applied as an algorithm, it needs of course an input, in the form of “atoms of computation” (Chomsky, 2010, p. 57): the lexical units. These units can be anything from words to affixes, from a phrase to a clause (Bickerton 2009b, p. 4; Berwick 2011, p. 87). All non-combined lexical units are stored in an inventory called the lexicon, from which Merge retrieves units to build sentences.

In Chomsky’s writings these lexical units are often conflated with concepts, although the equivalence is never made explicit, and the possibility that they may be different is acknowledged. Lexical units are alternately called “lexical items/concepts”, “conceptual atoms of thought” (Berwick & Chomsky, 2011, p. 39), or “internal conceptual symbols” (Chomsky, 2010, p. 57), suggesting an equivalence between lexical items and concepts. Other passages however suggest that there may be a difference since they are evoked independently: “words and concepts of human

language” (Chomsky 2010, p. 57; Berwick & Chomsky 2011, p. 39). Ultimately, Chomsky is silent on the relation between words and concepts, stating only that “the extent that they differ [...] is far from a simple question” (Chomsky, 2010, p. 57). Of course much work has been done on this question (see Meunier, 2006 for an overview of the philosophical tradition; Machery, 2009 for an overview and critique of the cognitive science approach), which unfortunately Chomsky does not mention.

One aspect of these lexical items Chomsky is clear about is their non-referentialist property. Chomsky has argued in many instances (Chomsky 2000a; 2005; 2010; Berwick & Chomsky 2011) that “even the simplest words and concepts of human language and thought lack the relation to mind-independent entities that appears to be characteristic of animal communication.” (Chomsky, 2010, p. 57) He recognizes that this is contra “Frege, Peirce, Tarski, Quine, and contemporary “externalist” philosophy of language and mind”, but nevertheless maintains that non-referentialism is one of the “critical differences between human conceptual systems and symbolic/representational systems of other animals,” (ibid., p. 57) and must be accounted for in language evolution.

Lexical items furthermore have lexical features which licence the Merge with other lexical items. In fact, lexical items are sometimes construed as no more than “a list of features” (Berwick, 2011, p. 84; see also Collins & Stabler, 2011, p. 2), which will interact in the generative procedure (also called a derivation), in a process called feature checking. Individual lexical items will thus have features such as $+f$ or $-f$, which can cancel each other, erasing the features in the course of the derivation. Other features can be represented as $=x$, which select a certain category of feature (Berwick, 2011, p. 87). Features can be semantic, syntactic or phonetic (Collins & Stabler, 2011, p. 2), with each playing differing roles in the derivation or at the interfaces of the CI and SM systems. How lexical items, lexical features and Merge interact is illustrated in the following section.

2.4. Application of Merge and lexical features

The building blocks of the language faculty proposed within the MP interact in specific ways in an attempt to account for the formation of all and only grammatical sentences. Berwick (2011) illustrates a derivation as follows, seen in context in figure 1: “For example, we take *the* to be a determiner, + *det*, selecting the feature *n*(oun),

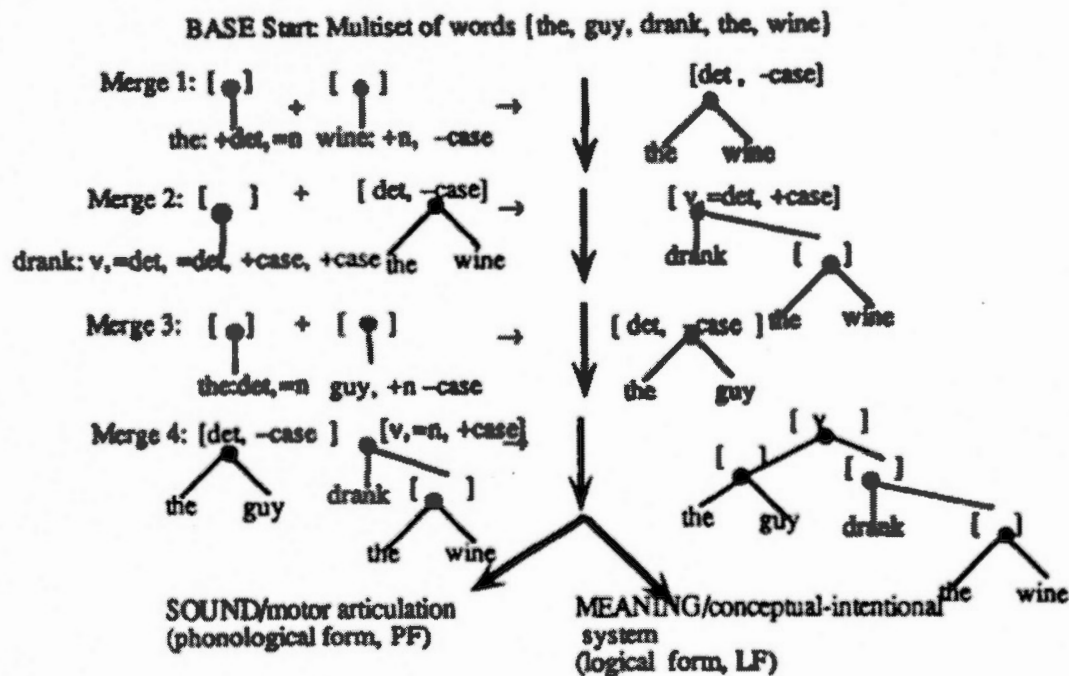


FIGURE 1 - (taken from Berwick, 2011, p. 86) Example of a derivation through four applications of Merge. Merge applies either to words from the lexicon or previously merged lexical units. Features get checked through each Merge, until all features have been eliminated, with the result sent to the “motor articulation” (i.e. the SM system) or the CI system.

so it has the formal features $+det, =n$; while *wine* is marked $+n, -Case$.⁷ The $=n$ feature can select the $+n$ feature, so Merge can apply in this case.” (Berwick, 2011, p. 87) In this example only syntactic features are shown. Features thus get “checked” as they are paired with the appropriate corresponding feature in a Merge operation. Whatever features cannot be checked in that application of Merge remain in the new lexical item, triggering further applications of Merge. The derivation continues until all features have been checked.

2.4.1. Asymmetry of interfaces

Once feature-checking is complete, the phrase structure is sent through the interfaces to the SM and CI systems. At this point the output is structured as the binary branching seen at the bottom of figure 1. At the interface to the SM system, it is serialized, placing one word after the other to allow for ‘externalization’ through speech or sign. It is suggested that linearization may be restricted to the SM system, with “no order [...] introduced in the rest of the syntactic computation” (Chomsky, 2005, p. 15). In other words, rather than linearizing the output, the CI system interprets the phrase as a branching tree, in all its hierarchical structural complexity, and attributes meaning. In a simple case such as figure 1 this seems trivial, but as is shown in the next paragraph, in cases where the structure is more complex or undergoes more radical transformations at the SM interface, the difference can be significant. This difference in the two interfaces has been called the ‘asymmetry of interfaces’, a purported fact which is thought to be related to language evolution, as will be shown in chapter III.

⁷ “Case” is a grammatical category which reflects the roles of nouns and pronouns within a sentence. Variations in case account for the difference between subject and object, which in turn accounts for (among other things) the difference between ‘I’ and ‘me’. Some languages across the world use case on other categories of words, and have multiple different cases.

How this asymmetry is significant becomes apparent in examples of the use of Internal Merge. Internal Merge is used in cases that were previously explained through the use of “Move”, a syntactic operation which was thought to be necessary to explain the formation of *wh*-questions. Consider for instance the interrogative *What will you say?* The derivation of this sentence (shown below) calls on applications of External and Internal Merge, as well as the lexical item with only the feature *Q*, “for “question,” that attracts *what*, while *what* possesses a $-Q$ feature” (Berwick, 2011, p. 89). As a word is recruited through Internal Merge, it occupies both its original position as well as the new position at the head of the phrase. The deletions, here represented as a strikethrough, are not thought to happen during the derivation itself but for communication. The following derivation is adapted from Rizzi (2009, p. 80):

(xiii) Derivation

<u>Lexical items</u>	<u>Type of Merge</u>
(a) {say what}	Ext Merge
(b) {you {say what}}	Ext Merge
(c) {will {you {say what}}}	Ext Merge
(d) {you {will { you {say what}}}}	Int Merge
(e) {Q {you {will { you {say what}}}}}	Ext Merge
(f) {will+Q {you { will { you {say what}}}}}	Int Merge
(g) {what {will+Q {you { will { you {say what }}}}}}	Int Merge

The deletions marked by a strikethrough are thought to happen only at the SM interface, where duplicates are removed in order to reduce computational load. Berwick & Chomsky (2011) note that “serial motor activity is computationally costly”, and that “with all but one of the occurrences of [duplicates] suppressed, the computational burden is greatly eased. The one occurrence that must be pronounced is the most prominent one, the last one created by Internal Merge.” (pp. 31-32)

Deleting duplicate words before externalization alleviates the computational load, but it does so at the expense of ease of interpretation. At the CI interface on the other hand, there is no serial motor activity involved, therefore there is no need for deletion: “the full internal expression is interpreted at the semantic [i.e. CI] interface” (Chomsky 2010, p. 54) Thus the CI system analyses the full structure with no lexical items suppressed. Consequently, suppressions are undone during parsing (interpretation) of communication, and the structure of the sentence is reconstructed for interpretation at the CI system (for a more complete understanding of the parsing operations, see Berwick, 2011, pp. 95-98).

2.4.2. Additional complexities in syntax

These two elements, Merge and lexical items/features, as well as the interfaces to the systems, are expected to account for all syntactic constructions in every language across the world. This includes more difficult syntactic phenomena two examples of which are Binding and Control (Bickerton, 2009b, p.8; the following summary and illustrations are taken from this article). Binding is the relationship between anaphors (himself, her, etc.) and the subjects to which they refer. Control is a way to describe the subject of certain verbs in subordinate clauses. Compare the following sentences:

(xiv) Mary told Bill to leave immediately.

(xv) Bill promised Mary to leave immediately.

The subject of “leave” is understood as being Bill in both cases, despite their inversion in the first clause, a fact which must be accounted for in any linguistic explanation. Linguists are very familiar with these issues as well as many others, as they are often the benchmark on which a theory can be tested (see Bickerton 2009b for a more in-depth description of this phenomena, as well as Tallerman *et al.*, 2009 for additional syntactic peculiarities to account for in any syntactic theory). Researchers within the MP are currently working to explain all these phenomena with

only Merge and lexical features, with some measure of success. As should now be clear, the gamble (as Boeckx, 2006 calls it) that the MP will be able to do so is far from won, with specific mechanisms and details of features and derivations still being debated.

It may well be that additional mechanisms are needed to account for all syntactic phenomena. At least three of those mechanisms may in fact already be implied in the interactions between Merge and the lexical features. The first is the “checking” of lexical features which seems to be stemming not from Merge, but from the features themselves. Just as the pieces of a puzzle fit together to eliminate odd-shaped ends, it could be argued that feature-checking comes ‘for free’ from the features themselves. The second mechanism that may or may not already be accounted for is how the features which are not checked are ‘projected’ to the Merged set, licencing further applications of Merge. This, again, does not seem to be a property attributable to Merge, but could perhaps be another property of the features. Chomsky (2008, 2013) has evoked a “labeling algorithm” (quoted in Collins, 2014) to account for the ascription of features to the Merged sets. The third is the relations between lexical units, and how those relations are a consequence of their relative position in the tree-structure. The fact that such relations exist is clear from the phenomena explained in the previous paragraph, such as control. According to Berwick (2011), such relations are explainable through the “temporal sequence of Merge,” or “derivational history” (p. 92), with no additional mechanism stipulated. Whether it is indeed the case that the sequence of Merges can explain all structural dependencies or not is an open question (Bickerton, 2009b). Of course, with the MP focusing on optimality, attempts are underway to reduce these—and indeed all—interactions to merely properties of Merge or of the features, without resorting to additional stipulations (see e.g. Collins, 2014).

2.5. The place of Merge in the faculty of language

According to the MP, Merge plays a central role in all syntactic processing, an idea which, as will be shown in chapter III, becomes a driving force behind Chomsky's account of the evolution of language. In anticipation of the arguments, this section evaluates just how important Merge is to the language faculty.

2.5.1. Computational load

Contrary to the P&P approach, the MP offloads the computational complexities of syntactic phenomena onto the lexical features. Whereas the P&P theories assumed that the syntactic algorithms were many and interacted in complex ways, the MP has reduced the algorithm to the utmost, to posit only Merge. However the cost of such apparent simplicity was to offload the computational complexity onto the lexical features. It is they which account for limitations in the production of sentences, distinguishing grammatical from ungrammatical sentences. Indeed without features, Merge alone does not yield well-formed sentences, having no constraints on applications of the combinatorial algorithm. Moreover, the features may even drive the entire syntactic process, as argued in the next sub-section.

2.5.2. Merge and recursion

For those familiar with previous versions of the generative grammar enterprise, it is interesting to note the MP uses a bottom-up approach to sentence-construction, rather than the previous top-down approach, with implications for recursion. In previous models, such as the one schematized in chapter I (section 1.4.1; recursion – definition 1: embedding), the sentence structure was built first using recursion-embedding, with

rewrite rules of the type “ $S \rightarrow NP VP$ ”, then populated with lexical units. In the MP, lexical units are the building blocks, with the structure appearing only through successive applications of Merge. As noted earlier, this approach shifts the type of recursion from embedding to cyclical. Yet recall that one of the essential features differentiating recursion from iteration is the hierarchical structure which emerges—a structure which may not be possible without lexical features.

Indeed Merge is not entirely responsible for the structure-building insofar as it is triggered by the lexical features, since it is they that necessitate further applications of the algorithm to complete the feature-checking. Collins & Stabler (2011, pp. 17-18) illustrate through various quotations from researchers in the MP that lexical features are understood to drive the applications of Merge; Berwick (2011) expresses this idea succinctly: “All structure-building is feature-driven” (p. 88). In other words, the successive applications of Merge are possible and necessary only insofar as the feature checking is incomplete. Bickerton (2009b) has noticed the same issue: “it is not simply Merge, but rather Merge + lexical material that constitute the recursive process, as well as force it to result in a hierarchical structure.” (p. 7, footnote 2) It is thus incomplete to affirm that Merge accounts for recursion in the language faculty, as it depends on lexical items and their features as well for its recursive generative process.

2.6. The broader scope of Merge

In light of the significant role of lexical features regarding important properties of the language faculty such as identification of grammatical sentences, as well as recursion itself, the question arises whether Merge on its own, now recognized essentially as a simple combinatorial algorithm, could be used in other cognitive processes. As will be shown in chapter III, Chomsky places Merge within FLN, hence believing that it is

a uniquely human and uniquely linguistic phenomenon. If it could be shown that it is found in other human cognitive capabilities, or even in other species, this would call into question its place within FLN.

2.6.1. In human cognition

Although not calling upon Merge itself, Corballis (2011) argues that recursion is not restricted to language processing, but also involved in other capabilities unique to humans. According to Corballis, the ability to project our plans into the future and reflect upon our memories “is recursive, in that one can insert previous personal experience [or future imagined experience] into present awareness,” (Corballis, 2011, p. 85) in what he dubs “mental time travel” (ibid., p. 81). It is also the foundation for our theory of mind, which “is recursive, in the sense that it involves the insertion of what you believe to be someone else’s state of mind into your own.” (ibid., p. 133) For Corballis then, recursion “is the primary characteristic that distinguishes the human mind from that of other animals.” (ibid., 2011, p. 1)

Could these processes be accounted for through applications of Merge? Insofar as applications of Merge are driven by feature-checking, which is what creates the hierarchical structure, the answer is clearly no. However there are similarities to the extent that mental time travel and theory of mind rely on recursion, and could conceivably be the result of a Merge-like operation which does not rely on features but rather on some other concept-matching properties. Interestingly, this is not problematic for Chomsky since, as will be explained in chapter III, he believes that the power of Merge and the language faculty apply to cognition more generally, enabling “thought and planning” (Chomsky, 2010, p. 55), and presumably the processes described by Corballis. However for Chomsky these capacities are dependent upon the language faculty, presumably obviating the need for the

postulation of either a non-linguistic Merge or non-linguistic features. The capacities would thus have appeared in conjunction with the language faculty. This is at the very least a debatable claim, which could be challenged by finding recursive cognitive processes in non-human animals.

2.6.2. In non-human cognition

Hauser, Chomsky and Fitch have already mentioned that recursion could play a role in certain areas of animal cognition such as navigation or social relations (HCF 2002, p.1578), but it was unclear what type of recursion was in question. Regarding Chomsky's more recent work, if Merge or a Merge-like operation does indeed play a broader role in human cognition, it would be reasonable to suppose that it is present in other species, implying that its origin may be traced back to a common ancestor between humans and those species. Bickerton (2009b) argues that if Merge is defined simply as a combinatorial algorithm which can apply to its own output (i.e. omitting the role played by features), then it is surely pervasive across species. The point is illustrated with the following example (Bickerton, 2009b, p. 6): "a bird building a nest, for example:

Step 1: Weave two twigs together.

Step 2: Interweave a third twig with the interwoven pair.

Step 3: Interweave a fourth twig with the interwoven three, etc."

Notice that this process can certainly be captured by Merge since the input of each step (except the first, of course) is the output of the previous. Thus if Merge truly is only such a simple algorithm, then it is surely widespread among all species who are able to act according to the result of their previous action.

One might argue however that the process of nest-building described above does not quite mirror the full power of Merge as it is used in linguistic processing. As

explained earlier, Merge operates on linguistic items repetitively until all features are checked, after which the output is shipped to the sensory-motor systems, which externalizes it. The process is therefore more complex than merely the ability to act upon one's previous output since it involves keeping in memory, or 'stacking' information that is created along the way and creating an internal structure where the dependencies are not simply the effect of linear order. In other words, once the phrase is externalized, the words produced will not always be dependent merely on the previous words uttered; the first word can be dependent upon later words, implying that there is a structural complexity not captured in the nest-building example.

But this type of structural dependence which relies on "stacking" is present in at least one other species. Caledonian crows are known for their cunning use of tools and multi-step planning. In a series of experiments Wimpenny et al. (2009) showed that some crows were able to master sequential tool use, meaning that they would effectively use a tool to make available a second tool which would in turn be used to gain access to food: "the most demanding condition requiring the use of three tools in a sequence" (Wimpenny et. al., 2009, p. 4). Just as is the case with linguistic use of Merge, this implies the stacking in memory of multiple items until the desired outcome is expected, after which execution is carried out:

Step 1: Plan on retrieving food

Step 2: Plan on retrieving tool A to [retrieve food.]

Step 3: Plan on retrieving tool B to [retrieve tool A to [retrieve food.]]

Step 4: Execute plan.

Because the plan in Step 1 is dependent upon execution of the plan in Step 2, and Step 2 is not desirable except insofar as it allows for Step 1, with the same relation following throughout the following steps, all steps must be formulated in a hierarchical structure before being successfully carried out. Only once the entire plan is formulated can it be shipped to the SM system, ultimately allowing for retrieval of food.

This is strikingly similar to the operation carried out by Merge, especially if one overlooks the role played by lexical features. Just as was mentioned regarding human cognition more generally, these examples cannot be the result of Merge itself, since it is thought to be restricted to linguistic cognition, and relies on features for its structure-building. However the example of tool retrieval of Caledonian crows does suggest that a Merge-like process is used, implying that Merge may be related to a more general combinatorial capacity perhaps not restricted to human cognition, much less human linguistic cognition, suggesting the possibility of common descent.

Of course such comparisons only *suggest* that Merge could stem from common descent. Similarities in traits across species can be the result of homologs or analogs (Stearns & Hoekstra, 2000). Whereas homologs are similar traits that are the result of common descent (p. 238), analogs (or homoplasy) are the result of convergent evolution, not shared genetic material (p. 236). To show that the capacities shared by New Caledonian crows and humans are the result of homology rather than analogy, research is required into such capacities in more closely related species, or into the genetic underpinnings of the capacity.

If it could be shown that Merge is not restricted to linguistic phenomena, it may in fact be a boon for the MP. After all, the aims of the biolinguistics aspect of the MP is to show that the mechanisms involved in language can be found in other aspects of biology or cognition. If Merge is a broader cognitive tool, “the hypothesized role of UG [Universal Grammar: the genetic endowment for language acquisition] declines, and the task of accounting for its evolution is correspondingly eased.” (Chomsky, 2010, p. 51) Notwithstanding, this is not the approach adopted by Chomsky in his attempts to account for language evolution, as will be seen in chapter III, and in fact would run counter to his hypothesis.

2.7. Conclusion

With the general claims and mechanisms of the MP laid out in this chapter, we are in a position to properly assess the arguments pertaining to language evolution brought forth by Chomsky in his writings subsequent to the articles co-authored with Hauser and Fitch. As will become clear, understanding the mechanisms proposed and the ways in which they interact is crucial to understanding how and where Chomsky's theory of language evolution fails.

CHAPTER III

MERGE AND SALTATIONISM

3.1. Chomsky's Minimalism and language evolution

Up to this point we have seen Chomsky's first systematic approach to language evolution in chapter I, in the article co-authored in 2002 with Hauser and Fitch. As argued earlier, the greatest problem with that publication was that the content of FLN, recursion, remained undefined and no specific mechanism was proposed to account for it. This is somewhat curious after having seen in chapter II how Chomsky does (and did) in fact have in hand the particular mechanisms within the Minimalist Program (henceforth MP) meant to account for language acquisition, production and comprehension⁸. As we will see in this chapter, not surprisingly, Chomsky has taken the findings of the MP and applied them to language evolution. The two most recent articles dealing explicitly and exclusively with the topic are used to summarize his views and arguments: Chomsky (2010) and Berwick & Chomsky (2011). Both articles advance many of the same claims and arguments, with certain additions or clarifications in one or the other, and are taken here as complementary and generally interchangeable.

As pointed out in the introduction, Chomsky's saltationist position has been criticized by many, typically through proposals of alternative explanations, or through differing interpretation of empirical data. Most of the criticisms concern the article coauthored

⁸ Although Hauser and Fitch may not have wanted to commit to the Minimalist Program.

in 2002 with Hauser and Fitch, but few have tackled his more recent contributions to the language evolution debates. In this chapter my aim is to show that even accepting at face value the findings of the MP, Chomsky's account fails to deliver on many points. Remaining within the MP has the advantage of not relying on a competing paradigm which depends on future research to be validated, nor on the interpretation of empirical data, which could turn out to be contentious.

To properly frame and critique Chomsky's saltationist approach, I review each of the arguments brought forth, then proceed to examine in detail whether they stand up to scrutiny. In so doing, I reveal many problems with Chomsky's arguments, particularly when he draws on other disciplines such as paleoanthropology or evolutionary biology. These issues cast doubt on the validity of his hypothesis, but nevertheless fail to falsify it. There is however one very significant point which is problematic not because of evidence garnered from other disciplines but because of findings from within the MP: lexical items and lexical features are virtually left out of the evolutionary account. Berwick and Chomsky (2011) in fact include lexical items in the proposed chronology of language evolution, all the while admitting ignorance regarding practically everything to do with them (Berwick & Chomsky, 2011, p. 40), as well as making no mention of lexical features. I argue that the treatment of lexical items amounts to more than a simple "gap" which needs to be filled: rather inclusion of lexical items and lexical features in the evolutionary account precludes Chomsky's saltationism.

In sections 3.2 and 3.3 I present a summary of the content of Chomsky's recent articles, laying down the premises for the arguments advanced by the authors, then the arguments themselves as well as the conclusions, and finally their proposed chronology of the evolution of the various parts of the language faculty. I follow with a mention of the positive aspects of Chomsky and Berwick's account of the evolution of language in section 3.4. Starting with sections 3.5 and 3.6 I critically address each

of the points advanced in their hypothesis, casting doubt on many of their assumptions and conclusions. Finally, section 3.7 tackles the evolution of lexical items and features, showing that Chomsky's account ignores the interacting mechanisms of his own MP theories, thus not only falling short of a complete account but also arriving at erroneous conclusions.

3.2. Premises

3.2.1. Setting the stage

Chomsky has a way of framing problems in linguistics and cognitive science such that they fall squarely within his own paradigm, to the exclusion of all others. Language evolution is no exception, as can be seen from the handful of articles he has written or co-authored on the topic. In these articles he recasts the problems of language evolution in light of the MP, claiming that "for those concerned with evolution of language, the minimalist program must surely be a central concern" (Chomsky, 2010, p. 52). Indeed for Chomsky the MP is not to be taken as a proposal regarding the mechanisms of the language faculty, but rather simply the impetus to reduce or simplify them, regardless of one's theoretical commitments. It is thus claimed to be "largely theory-neutral": "whatever one's beliefs about design of language may be, the questions of the minimalist research program arise" (ibid., pp. 51-52). Chomsky thus takes for granted that "serious" (his words; ibid., p. 51) research into language evolution must be done through the lens of the MP.

If the MP were indeed theory-neutral, the above statements would be fairly innocuous, and simply an encouragement to researchers to curb the proliferation of proposed mechanisms involved in the language faculty, in the hopes of a better reconciliation with biology. But, as seen in chapter II, the MP does not in fact limit

itself to broad and abstract ideals regarding language study, but rather to specific hypotheses about the “recursive operations that enter into the computational system of language.” (ibid., p. 52). As is evident from even that short phrase, many assumptions are already packed into the approach, namely that language is a computational system, and that it involves recursive operations. Furthermore, according to researchers from within the MP, that recursive operation should “approximate the simplest form possible, perhaps even reaching its limit” (Berwick & Chomsky, 2011, p. 52). Their proposed solution is the operation Merge, as seen in detail in chapter II.

Chomsky’s approach to language evolution is thus to take for granted that the proposals of the MP will someday be vindicated; and that in the meantime the hypothesis that Merge is the sole syntactic operation should be taken seriously. Furthermore, according to Chomsky all other approaches are problematic since “any stipulated device beyond Merge carries a burden of proof: the complication of UG [Universal Grammar: the genetic endowment for language acquisition (p.51)] must be based on empirical evidence.” (Chomsky, 2010, p. 52) This evidence is of course understood to be non-existent. His arguments therefore assume that Merge is the sole component to be accounted for, explaining—as we shall see—how he can advance a saltationist position with respect to the appearance of language.

3.2.2. Human revolution and “Out of Africa”

The first premise for the saltationist approach is that language appeared fairly recently in evolutionary time, and apparently very quickly. Evoking the archaeological record, both articles claim that a “fairly general consensus” among researchers (Berwick & Chomsky, 2011, p. 19; a similar phrase is found in Chomsky, 2010, p. 58) places the appearance of language sometime between 100 000 and

50 000 years ago. This coincides with the “great leap forward”, or “human revolution”: an explosion of human technological and cultural innovation apparent in the archaeological record, coupled with the migration of *Homo sapiens* out of Africa. The assumption behind this premise is that there must have been a genetic change in the species which would explain this sudden difference in behaviour, which is attributed to the appearance of language.

3.2.3. Species-wide language faculty

A second premise for Chomsky’s and Berwick’s arguments, which is in fact twofold, is that the faculty of language is species-specific and species-wide, and that it has remained essentially unchanged since its appearance. They point out that a human infant from any part of the world can be raised in a linguistic community in any other part of the world and acquire the language in the host community, implying that all humans have the same capacity for language-learning (Chomsky 2010, p.58; Berwick & Chomsky 2011, p.19-20). Thus there exists a species-specific capacity for language, implying that it is, at least in some measure, grounded in biology. Since it is also species-wide, it is also reasonable to assume that the heritable biological component stems from a common ancestor. It would have furthermore remained unchanged since it appears to be uniform across the species.

3.2.4. Language not for communication

A third and somewhat more contentious premise is that language’s primary function is—or was—not communication, but rather thought itself. Under this approach, communication, dubbed in these articles as “externalization”, happened afterwards as a way of externalizing thought. Berwick and Chomsky of course realise “how

difficult it can be to assign a unique function to an organ or trait” (Berwick & Chomsky, 2011, p. 25), but nevertheless attempt to demonstrate through three arguments that the language faculty was never designed for communication.

1) Primary use in thinking: Their first claim is that the current use of the language faculty is typically for thought. They point out that “statistically speaking, for whatever that is worth, the overwhelming use of language is internal—for thought.” (Berwick & Chomsky, 2011, pp. 25-26) Since, according to Chomsky and Berwick, language is more often used as a tool for thought than for communication, it stands to reason that its primary function was to organise thought itself.

2) Modality-independence: Another reason taken to suggest that externalization is a secondary function of language is that it does not depend on a single modality. Recent work has shown that the full depth and capacity of spoken language is found in sign language as well: the structural properties, the neural localization and the manner of acquisition all are very similar to spoken language (Chomsky, 2010, p. 56; Berwick & Chomsky, 2011, p. 32). The authors take this to imply that communication is an accidental aspect of the language faculty: since the language faculty is able to recruit various sensory-motor systems for communication, its primary function must lay elsewhere.

3) Asymmetry of interfaces: Chomsky’s last argument for communication not being the primary function of language is that whenever there is a conflict between communicative efficiency and (cognitive) computational efficiency, the latter will win. As seen in both the 2002 article with Hauser and Fitch and in the MP more generally, Chomsky proposes that the syntactic component is central to the language faculty, and interfaces with the Conceptual-Intentional (CI) and Sensory-Motor (SM) systems. However in these more recent articles it is assumed that there is an asymmetry in the interfaces, with the CI system being privileged over SM system

(Chomsky, 2010, pp. 54-55). As has been explained in chapter II, when the syntactic system interfaces with the SM system, certain lexical items which occur more than once will be suppressed, despite their use at the CI interface. This is thought to be because “serial motor activity is computationally costly” (Berwick & Chomsky, 2011, p. 31), and therefore “computational efficiency yields the universally attested facts: only the hierarchically most prominent position is pronounced.” (Chomsky, 2010, p. 55) This allows for a lightening of the computational load when it comes to the SM interface, but yields ambiguities in interpretation. Computational efficiency is thus done at the expense of communicative efficiency. At the CI interface however, all lexical units are assumed to be present to allow for interpretation of the sentence. Thus, since computational efficiency seems to have priority over externalization at the SM interface, and no such compromise exists at the CI interface, Berwick and Chomsky take this to mean that external communication is an ancillary function of language, whereas the mental aspect would be the ‘true’ and original function of language.

3.2.5. The atoms of computation

The final premise regards the nature of the “atoms of computation” on which Merge operates. Described in chapter II under the rubric “lexical items” (section 2.3.2), their description is somewhat vague: they are something akin to concepts, perhaps even finding their roots in “conceptual structures [...] found in other primates [such as] actor-action-goal schemata, categorization, possibly the singular-plural distinction, and others.” (Berwick & Chomsky, 2011, p. 39) Despite these tentative parallels, Berwick and Chomsky maintain that “the atoms of computation, lexical items/concepts, appear to be uniquely human.” (ibid., p.39) The reason given for their uniqueness is their property of non-referentialism, discussed at greater length in chapter II. “These properties of lexical items [*viz.* those which are a consequence of

non-referentialism] seem to be unique to human language and thought and have to be accounted for somehow in the study of their evolution.” (ibid., p. 40) Unfortunately, the authors do not seem to have any leads on how to address this, and enjoin: “How, no one has any idea.”

3.3. Chomsky’s conclusions

Chomsky and Berwick gather these threads together to propose what they believe to be the “simplest” explanation for language evolution. According to their assumptions, the faculty of language would have appeared suddenly, sometime between 50 000 to 100 000 years ago, with a change that would have marked a transition from having no linguistic competence to having full linguistic competence, essentially unchanged from our contemporary language faculty. From there they go on to claim that “the simplest assumption, hence the one we adopt unless counter-evidence appears, is that the generative procedure emerged suddenly as the result of a minor mutation.” (Berwick & Chomsky, 2011, p. 29; cf. Chomsky 2010, p.55). This “minor mutation” would have led to great adaptability in the humans which possessed it, gaining great advantage with respect to natural selection, accounting for both the “great leap forward” and the emigration out of Africa.

3.3.1. Merge as keystone

That “minor mutation” yields the operation Merge. Because language is assumed to have appeared suddenly, “we would expect the generative procedure to be very simple” (Berwick & Chomsky, 2011, p. 29). This simplicity is what allows the claim that the language faculty is able to emerge in a single step: if the operation is so simple that it cannot be broken down into component parts, or sub-routines, it follows

that it can only have appeared fully-formed. Merge, being “the simplest mode of recursive generation” (Berwick & Chomsky, 2011, p. 30), fits the role perfectly. This is at the heart of the saltationist argument: Merge is so simple that it cannot logically have appeared through successive minor mutations, and therefore would have emerged in a single step, yielding qualitatively huge differences in the species.

Despite its simplicity Merge is believed by Chomsky and Berwick to have precipitated such a great change because it would act as a keystone, drawing all other requisite cognitive and physiological mechanisms together to allow for language. Implied is the thought that the interfaces to the CI and SM systems come ‘for free’, requiring no supplemental cognitive development. It is unclear in either article (Chomsky, 2010; Berwick & Chomsky, 2011) whether Merge should be understood as truly appearing at this time, or as being the result of an exaptation taken from a more general combinatorial algorithm previously present but not applied to language. In either case, Merge’s keystone role draws on the distinction laid out in Hauser, Chomsky & Fitch (2002) (see chapter I for clarifications): in this case Merge (as well as its interfaces to the CI and SM systems) is the only component of FLN and all other components necessary to language are part of FLB, thus already present when Merge appeared⁹.

⁹ Chomsky has subsequently suggested (p.c. 2013) that lexical items as well may need to be attributed to FLN. But if this were the case, then the whole saltationist account as presented by Chomsky and Berwick falls apart since it hinges on the fact that Merge, a simple algorithm, could have appeared in a single step. Additional elements in FLN indicate that a gradualist account is far more plausible. For a complete discussion of this and related topics, see section 3.7.

3.3.2. Communication's non-adaptationist origins

Since the initial function of the faculty of language would not have been communication, but “an instrument of internal thought” (Berwick & Chomsky, 2011, p. 32), Merge would not immediately lead to communication. Instead, it would enable planning, reasoning, combination of symbols, etc. This would confer a selectional advantage on the individuals who possess the trait, which would therefore spread (Chomsky, 2010, p. 59). Only later (at some unspecified time) would the sensory-motor system be recruited to start externalizing these thoughts, which again would lead to a selectional advantage. Note however that in this scenario the entire language faculty is not the fruit of selection for communication, but rather appeared fully formed thanks to a minor mutation, which only incidentally—and later—allowed for communication.

According to Berwick and Chomsky, there is furthermore good reason to believe that the communication that stems from the use of the language faculty is not only secondary, but not an adaptation. Since spoken languages vary widely from one population to another subject to the vagaries of cultural change, “it follows that externalization may not have evolved at all” (Berwick & Chomsky, 2011, p. 38), explaining the lack of species-wide uniformity. It is supposed that the “problem of externalization”—that is, the particulars of the mapping of Merge to the SM system—would have happened once the language faculty was in place, “a problem addressed by existing cognitive processes, in different ways, and at different times.” (ibid., p. 38) The development of language for communication would therefore imply no biological change—and consequently no evolutionary change—in the organisms.

3.3.3. No proto-language

According to Chomsky and Berwick, this account precludes the possibility of any sort of proto-language. Language would have appeared with Merge, whose simplicity precludes its appearance through successive steps. As a result, there could not have been a simpler precursor to Merge which could have been added onto or complexified in order to yield Merge. According to Chomsky (2010), there nevertheless “are many proposals involving precursors with a stipulated bound on Merge” (p.53) which would allow for the combination of only two words, then later in evolutionary time three words, then four, etc. But “there is no rationale for postulation of such a system: to go from seven-word sentences to the discrete infinity of human language requires the same recursive procedure as to go from zero to infinity” (Berwick & Chomsky, 2011, p. 31). There is furthermore “no empirical evidence from the historical or archaeological record for such stipulations” (Chomsky, 2010, p. 53). Since language is defined through the use of Merge, and having Merge is a binary (you have it or you don’t), it follows that no additional steps can be added without arbitrary complications of the evolutionary account.

3.3.4. Chronology of language evolution

Having reached the conclusions outlined above, Berwick and Chomsky propose a sequence for when these various elements would have evolved in a broad chronology of language evolution. The first element they assume to have appeared is the lexical items: “In some completely unknown way, our ancestors developed human concepts.” (Berwick & Chomsky, 2011, p. 40) Repeating the ignorance admitted to earlier regarding lexical items (and attributed to the entire body of research), no specific timeline or trigger is given for this element of the language faculty.

The next element which would have appeared is Merge. It is assumed to have happened “at some time in the very recent past, maybe about 75 000 years ago” (Berwick & Chomsky, 2011, p. 40). As explained earlier, a slight rewiring of the brain would have taken place, allowing the use of Merge on the previously developed concepts.

The final stage is the recruitment of the sensory-motor system for the use of spoken or gestured language. Berwick and Chomsky explain: “At some later stage, the internal language of thought was connected to the sensory-motor system, a complex task that can be solved in many different ways and at different times, and quite possibly a task that involves no evolution at all.” (2011, p. 41) This satisfies the assumption that the language faculty was not primarily for communication, and is thought to account for the variety of spoken and signed languages.

This completes the summary of Chomsky and Berwick’s account of the evolution of the language faculty, broken down into premises and conclusions. This structure allows for a clear picture of the theory, which in turn facilitates the critique which will follow.

3.4. Strengths of the theory

Before picking apart the account above, it is worth pointing out that despite its shortcomings, it does have many positive aspects. One thing that language evolution theories must account for is the selection pressures that would have led to the development of the language faculty. Why would *Homo sapiens*, of all primates, mammals or even species, be the only species with such a complex and explicit communication system (or so it seems)? There must have been selection pressures at work which would not have been present for other species. This is a difficult problem

for all language evolution theories (and evolutionary psychology in general) since, as Durrant & Haig (2001) point out “we are barred from access to the appropriate detailed information about ancestral environments [such as] relevant data concerning genetic differences, reproductive success, and so on” (Durrant & Haig, 2001, p. 359). Since the environment of evolutionary adaptedness relating to language evolution is long gone, researchers are at pains to provide convincing evidence regarding the selection pressures proposed in their models (see Laland & Brown, 2011, pp. 124-128 for a critical discussion of this problem).

Chomsky’s model partly succeeds in sidestepping this issue insofar as it does not require feedback from the environment for the *shaping* of Merge since it appeared in a single step. Of course selection pressures were needed to maintain and spread the trait in the population, but recall that Merge’s appearance is assumed to have allowed for a full language faculty overnight. In such a case, it is easy to imagine that the “capacities for complex thought, planning, interpretation, and so on” (Chomsky, 2010, p. 59) that came along with it would have bestowed a great selectional advantage on the individuals and groups in which it was present. Their increased survival is thus virtually guaranteed in practically any environment, as testified by “Out of Africa” and the spread of humans all over the globe. Chomsky’s theory consequently practically obviates the need for the identification of specific selection pressures associated with language. Notwithstanding, it only partly succeeds in sidestepping the issue since all aspects of FLB which precede the appearance of Merge must still be accounted for by identification of the functions and selection pressures, thereby also identifying how they could have been present for (presumably) some time before being recruited to the language faculty. These aspects are not addressed by Chomsky or Berwick.

The most positive aspect of Chomsky’s account is that it proposes very specific mechanisms meant to explain the operations of the language faculty, and attempts to

account for them through evolutionary biology. As Steels (2000) puts it: “only clear scientific models that explain how language evolved (as opposed to enumerating conditions why language evolved) can be expected to steer us away from the many speculations that made the field suspect for a long time.” (p. 18) By proposing that language is produced through Merge and the interfaces to the SM and CI systems, Chomsky provides a strong and ostensibly falsifiable hypothesis regarding the language faculty. This in turn narrows significantly the possible evolutionary accounts for the appearance of language, providing the basis for a clear and precise history of its biological evolution. Note that this was the main criticism levelled in chapter I at the Hauser, Chomsky & Fitch (2002) article: their proposal that the property of “recursion” was the sole component of FLN fell short of providing a complete and clear hypothesis for language evolution since it failed to actually propose a specific mechanism. By approaching the issue through the MP, Chomsky solves this problem, although by the same token opens up his account to further criticism.

3.5. Critiques of premises

Many of the points made in sections 3.2 and 3.3 can be called into question, at times because the research in relevant disciplines is not so clear-cut, at times because there are logical leaps in the conclusions, or apparent misunderstandings with respect to the principles of evolutionary biology. Each of the sub-sections below refers to a corresponding subsection above, and suggests, where possible, why each of the premises or conclusions may be problematic.

3.5.1. The Minimalist Program at center stage

As mentioned in chapter II (section 2.2.2), notwithstanding Chomsky's claims, many researchers have rejected the MP, implying that it is far from theory neutral. One need only look to the work of Ray Jackendoff (2002, 2010), Philip Liberman (2002), Paul Postal (2004), and dozens of other prominent researchers to see that the MP is not necessarily the leading paradigm in linguistics, much less the only possible approach. The relevance of Chomsky's framing of language evolution exclusively in terms of the MP is therefore done at the exclusion of many other approaches, which may be in a position to propose alternative hypotheses regarding the evolution of the language faculty (for more on this see Jackendoff, 2010: "Your theory of language evolution depends on your theory of language"—but see also Clark, 2013, for a rebuttal). Nevertheless, beginning with specific (as yet non-falsified) hypotheses about the content of the language faculty is a good start for research into language evolution.

3.5.2. Human revolution

Much of Chomsky's argument centers on a very fast development of human cognition around 75 000 years ago, purportedly explainable only through the acquisition of language in humans. In the 1980's, archaeologists did indeed coin the term "human revolution" to indicate the apparent discontinuity brought about by a sudden increase in technology and culture some 40 000 years ago (Mellars & Stringer, 1989). However, as Hoffecker (2007) points out, "the extent to which the apparent discontinuity in the archaeological record at this time—as modern humans move their act from Africa to Europe—is due to contrast in archaeological visibility between the two continents (e.g. quantity of excavated sites, caves versus open-air localities) is still unclear." (p. 376) Other researchers (e.g. McBearty & Brooks 2000)

propose that there never was a “great leap”, and that the features attributed to the so-called revolution are found in the records of the African Middle Stone Age, tens or even hundreds of thousands of years earlier. Additionally, according to their account, not only did these elements appear earlier, but also more gradually, as opposed to in a single leap.

Furthermore, the “human revolution” may not be relevant to a saltationist argument since cognitive and technological developments do not always go hand in hand. Behme (2014) in a review of a recent publication by Chomsky (2012) argues that

it is not entirely clear that a detectable change in technology is a reliable indicator for an increase in overall intelligence and/or the arrival of linguistic abilities. By analogy, comparing the “archaeological record” of human technology of the 17th and 20th century a scientist of the 44th century might conclude that our species underwent a dramatic increase in intelligence during this time period. (p.15)

Interestingly, this argument rests on Chomsky’s own competence/performance distinction: the *capacity* for technological innovation (competence) and the actual innovations themselves (performance) do not necessarily follow one another closely. Moreover, there could have been a biological or cognitive change other than the appearance of language that explains the human revolution, such as finer motor skills for tool-making or cognitive developments for non-language-specific symbolic representation. There are thus reasons to question that the “great leap forward” was truly the consequence of a sudden biological change to the faculty of language, rather than some other biological transformation or change through accumulation of cultural knowledge.

Since the “human revolution” plays such a major role in Chomsky’s saltationist argument, its falsification would be a very significant blow to his position. Note however that it would not result in an invalidation of the saltationist argument, but merely a removal of the impetus for its supposition. Yet despite the arguments

outlined in this section, it is fair to say that most researchers in the fields of archaeology and paleoanthropology do believe that the “human revolution” happened at some point in the past, and link it to cognitive sophistication and language development (see McBearty & Brooks 2000, p.453-454, 486 for an overview of the literature). Chomsky and Berwick’s premise, though not as self-evident as it may seem, is nevertheless fairly reasonable.

3.5.3. Species-wide language faculty

This is probably the least contentious premise as it claims very little, namely that humans have a shared biological capacity to learn language. Note that it is not necessary to assume anything in particular about the biological capacity: arguments about its nature, size, specificity, etc. are not affected by the acceptance that there is a species-wide capacity for language. The only implication of this premise is that there is a biological capacity for language, and it is therefore reasonable to assume that it results from a common ancestor.

3.5.4. Language not for communication

1) Primary use in thinking: It is advanced that language is more often used for thought than it is for communication, implying that thought is the primary function of the language faculty. This argument suffers from two setbacks, the first of which is that it apparently relies entirely on intuition rather than data: no research is referenced to back up this fact (cf. similar unsubstantiated claims in Chomsky 2010, p.55, Chomsky 2012, pp. 11-12, Chomsky 2013). Nevertheless, it does sound intuitively plausible, although the particulars would need to be spelled out in order to demonstrate it. For instance, is thought truly, or always, language-based, as opposed

to image-based, or based in some other medium? And if internal thought is 'in a language' (e.g. English, French, etc.), then what would it have been like before externalized language existed (Jackendoff 2011, p.613)? Viger (2007) even proposes that the language of thought is acquired through public speech, in which case even if current use of language is mostly thought, that language would be impossible without the communication from which it stems. The claim therefore is not only unsupported, but runs up against contradictory intuitions, calling into question its legitimacy.

The second important setback to this argument is that even if the claim were empirically demonstrated, current function need not reflect the original function (cf. Laland & Brown, 2011, p. 92). This is obvious when one thinks of exaptations, where the current function is by definition not the original function. Berwick and Chomsky however seem to realize how weak the argument is, couching it as they do in dismissive terms: "for whatever it's worth..."

2) Modality-independence: Chomsky and Berwick argue that the fact that the language faculty can recruit speech *or* sign for communication is an indication that communication is a secondary aspect. The fact that sign languages as we know them are a relatively recent invention should not detract from this argument since there is a "long but checkered history" (Corballis, 2011, p. 57) of theories whereby language evolved from manual gestures, with speech arriving only later (see Corballis, 2011; pp. 57-59 for a short history). The argument is however otherwise quite weak since the complete opposite point could be made with the same premise: since the language faculty is able to use any modality for communication, it could be said to be perfectly suited for communication, able to use whatever means at its disposal to do so. This could imply that it *is* an adaptation for communication. Evidently, more research is needed before the modality-independence of language is brought to bear on this argument.

3) Asymmetry of interfaces: Finally, Chomsky and Berwick argue that since the language faculty is more efficient at the interface to the CI system than it is to the SM system, it follows that thought is a primary function of the language faculty, and communication a secondary function. Much, if not all, of this argument depends on acceptance of the research coming out of the MP, including the CI and SM systems, the nature of the interfaces, as well as particulars of sentence generation outlined in chapter II. As was mentioned earlier, many researchers have adopted other ways of studying language which have led them to posit far different mechanisms and cognitive architectures. Notwithstanding, from within the MP, the argument does appear to be well-formulated.

Nevertheless, talk of language's "primary function" does not seem to mesh with certain principles of evolutionary biology and even with Chomsky's saltationist approach more generally. The definition of a "function" has been hotly debated in biology and philosophy of biology throughout the 20th century. The emerging consensus was that a function is an effect which has been selected for (Wright 1976, Millikan 1984, Kitcher 1993). This in turn highlights the difference between an adaptation and a character which is merely adaptive. As Laland & Brown (2011) explain, an adaptation has been "favoured by natural selection for its effectiveness in a particular role" (p.91), whereas a character is adaptive if it currently increases reproductive success (note that a trait can also be both or neither).

If Merge (as a trait) remained and spread, then there must have been some positive selection pressure on it, therefore it must have a function. However because of the saltationist scenario, its current efficiency at various tasks tells us nothing about the selection pressures leading to its proliferation. Since Merge and the CI and SM systems are claimed not to have evolved since Merge appeared, it follows that their current state is identical to their state at the time when Merge first appeared. Yet for it to even be possible that the relative efficiency at the interfaces be an indicator of

original function, Chomsky and Berwick would need to concede that natural selection played a role in *increasing* the efficiency at the CI interface relative to the SM interface. In other words, that greater efficiency would need to be an adaptation, not merely adaptive. However their saltationist account is also non-adaptationist, admitting of no further evolution after Merge's appearance, thus precluding the possibility that selection pressures had an effect on the shaping of the faculty of language. This implies that whatever relative efficiency is present cannot be the result of the tinkering of evolution, and therefore cannot be an indicator of primary function.

To demonstrate that thought was the primary function of the faculty of language in Chomsky's scenario, one would need to show empirical evidence, presumably archaeological, that initial selection pressures were for the capacity for thought as opposed to communication, despite the fact that both capacities were available from the outset. No archaeological evidence of the sort has been proposed in Chomsky's writings, and indeed it is hard to imagine what that evidence would look like. The fact that the language faculty would be more efficient at the CI interface is therefore simply a matter of happenstance, and suggests nothing regarding its original role.

Besides, the fact that there are constraints that exist on the SM system which do not apply to the CI system may be related to the particulars of the systems themselves rather than computational efficiency constraints. For instance, the SM system seems to require serialization of linguistic output, even in the case of sign languages (which could in principle communicate in parallel using both hands) (Berwick & Chomsky, 2011, p. 32), whereas the CI system may not.

All the arguments advanced to support the premise that language is or was not primarily for communication are thus problematic. Nevertheless, it has not been demonstrated that language's primary use *was* communication, nor that it is

impossible for language's primary use to have been for thought. These arguments thus leave the debate still open, with consequences that shall become apparent in section 3.6.2: Communication's non-adaptationist origins.

3.5.5. The atoms of computation

The main criticism regarding this premise is of course just how vague it is. The conflation of words, concepts and animal cognition needs to be justified, as well as the changes from "conceptual structures" of other species to those we humans currently possess. These issues, as well as how the lexical items fit into the proposed chronology of language evolution will be addressed at greater length in section 3.7.

3.5.6. Premises: summary

In sum, of all the premises proposed by Chomsky and Berwick, the least contentious is the fact that humans possess a species-wide and species-specific language faculty, a premise accepted by most researchers in the field. The human revolution, the main impetus for the saltationist claim, is contentious within the fields of paleoanthropology and archaeology, but has nevertheless not been discounted. None of the remaining premises, be it the validity of the MP, the secondary role of communication, or the atoms of computation, have been demonstrated and remain open for debate, or in need of further research.

3.6. Chomsky's conclusions revisited

Chomsky and Berwick present what they believe to be the simplest assumption for language evolution, namely that language appeared in a single step, with the appearance of Merge. As mentioned in chapter I (section 1.1.3), saltations are thought to be the result of small genotypic alterations that confer a large discontinuous change in the phenotype. If that is the case, then language's appearance in a single step can be the simplest assumption only insofar as it can plausibly be related to a small change in genotype. Furthermore if saltations are considered rare, as some researchers maintain, then any claim to a saltation must be justified by showing that alternative gradualist scenarios fail to account for the trait. One way of satisfying both these criteria without relying on genetic analysis is to argue as Gould (1980) did that the given trait admits of no intermediate steps, as is the case with the movable joint of boid snakes (see section 1.1.3). The implicit justification behind this approach is that the genotypic change need not be specified if the phenotypic change cannot be divided into multiple steps. The idea is that no matter how many genotypic changes were actually necessary to produce Merge, at some point one of those changes (the last one) produced the change in the phenotype that led to the appearance of Merge where there was no Merge before. In other words, the argument requires there to be no genotypic change that could have led to any part of Merge without all of Merge. Chomsky and Berwick's saltationist argument needs therefore to be justified through evidence of a large discontinuous change in the species, and the fact that the trait(s) responsible for that change admit of no intermediate steps.

The principal reason Chomsky and Berwick assume that it is necessary to advance the saltationist claim is the acceptance of the "human revolution" theory. By their lights, since there is a large discontinuous change to account for, the proposal of a saltation to account for it is warranted. But how fast must change be in the archaeological record for it to be explainable only through saltations? According to Laland & Brown

(2011), research in the 1990's showed that "biological selection may be extremely fast, with significant genetic and phenotypic changes sometimes observed in just a handful of generations" (p. 133). In terms of the archaeological record, a swift change may be perceived as a sudden break when in fact it is a gradual—albeit very fast—change. Nevertheless, as discussed in section 3.5.2, although the existence of the "human revolution" is not without controversy, there is no knock-down evidence or argument against either the claim that it happened, or that it is the result of a saltation affecting language. Let us therefore assume for the sake of argument that it stands and turn instead towards other more clearly problematic parts of the argument.

Berwick and Chomsky claim that saltationism is the "simplest" conclusion, but it is "simple" only insofar as it is taken for granted that the language faculty hinges upon a single extremely simple algorithm, as opposed to a complex network of interacting parts. Compare for instance the fictitious and obviously false claim that a bird's wings appeared suddenly rather than through multiple successive steps, since it is a simpler assumption. Indeed it is simpler insofar as it reduces the number of steps needed to explain its evolution, but it is also incredibly implausible and therefore considered absurd. The reason why it seems impossible is because of the complexity of a bird's wing, involving as it does not only bones, muscles, nerves, veins, feathers, etc. but also a very specific architecture of moving parts in order to allow for flight, not to mention the brain structures dedicated to its control. This is why complexity is generally taken as a measure of plausibility of a gradualist evolutionary account; Pinker and Bloom (1990) observe that "the only successful account of the origin of complex biological structure is the theory of natural selection" (p. 710) and that the language faculty is just one of those complex structures. The question to be asking regarding the saltationist position is consequently to what extent language is simple or complex. If indeed it is extremely simple, then one might be able to argue that it appeared in a single step. Note however that the saltationist claim in this case would not be warranted so much because of "simplicity" or theoretical parsimony, but rather

because the mechanism is so simple that it *could not* have evolved in successive steps. This is precisely what Chomsky argues, as is clear from his conclusions regarding the inexistence of proto-language (see sections 3.5.3 and 3.6.3).

The simplicity claim regarding saltational evolution therefore rests on the simplicity of the language faculty. Incidentally, the MP's whole purpose is to reduce the mechanisms of the language to a minimum, so it is not surprising that the application of MP research to language evolution would lead them to the conclusion that the language faculty evolved through very few and very simple steps, perhaps even a single one. There is thus a certain circularity in the argument: the simplest assumption is that language evolved in a single step, but only because it is assumed that the language faculty is very simple, an assumption meant to help explain its biology, and therefore its evolution. The assertions regarding the language faculty and its evolution therefore both rest on the same unverified assumption of the desirability of "simplicity" and each plays a role in validating the other.

3.6.1. Merge as keystone

The circularity of the argument becomes more apparent at this point, when Berwick and Chomsky assert that it is *because* language appeared so suddenly that the generative procedure must be simple (Berwick & Chomsky, 2011, p. 29). As argued in the previous sub-section, the saltation claim itself rests on the assumption of a simple generative procedure, without which such a sudden change would be too implausible. But the language evolution theory is ostensibly based on the understanding of the language faculty, as testified by Chomsky's writing (2010): "Study of evolution of some system is feasible only to the extent that its nature is understood. [...] Accordingly, a sensible approach is to begin with properties of language which are understood with some confidence" (p. 45). This approach is later

termed “the only reasonable approach for inquiry into the evolution of any system.” (p. 52) In contrast, by arguing that the saltationist claim warrants the expectation that the language faculty arose thanks to a single simple operation, Chomsky flips the argument on its head, putting the weight of the argument on the evolutionary history rather than knowledge of the mechanisms involved. To a certain extent this contradicts his own “reasonable approach”; although to be clear, it does not invalidate his conclusions. Indeed it could be argued that the human revolution justifies the saltationist claim, which in turn substantiates the postulation of the simple recursive algorithm. However if the whole argument rests on the shaky foundation of the human revolution along with the assumption that it results from a biological change, then the argument erected upon it seems more and more precarious.

Perhaps the greatest difficulty in explaining the keystone role for Merge is accounting for the evolution of the mechanisms of FLB. If Merge is the only mechanism in FLN, and its appearance signalled the end of the evolution of the language faculty, then all the other components included in FLB must have evolved previously. This implies that mechanisms such as the CI and SM systems, lexical items, lexical features, phonology, semantics, etc. are either adaptations for some other unspecified function other than language, or the result of the interfaces to the CI and SM systems. Chomsky and Berwick are mute on how this could be possible.

3.6.2. Communication's non-adaptationist origins

Considering that the premises on which this conclusion rests have been shown to be at the very least problematic (section 3.5.4), the fact that language's communicative aspect is not an adaptation is far from demonstrated. Yet if the premises were abandoned, it may seem at first glance that the only requirement would be to drop the conclusion that communication does not have an adaptationist origin. In other words,

the conclusion does not seem to convey much more than the premises do, and they seem to relate only to one another.

In fact, much is at stake: communication must be a secondary, non-adaptationist by-product of the evolution of the language faculty if the saltationist argument is to hold. Consider for instance how a gradualist account of language evolution could propose that elementary aspects of communication appeared early (in one way or another), with further aspects or mechanisms spreading through selection pressures on communication, leading to an adaptationist account of communication and the language faculty (see e.g. Bickerton, 2009a). By contrast, Chomsky cannot assume a gradual development of communication, because it is antithetical to the saltationist position: if communication was fashioned through selection pressures and successive adaptations, then the language faculty itself would have been shaped by selection, implying that the saltationist position is impossible. Thus he could accept an adaptationist history to communication only at the expense of his saltationist claim. Of course, he could argue that any communication prior to the development of the complete language faculty is simply not language by definition: non-linguistic communication evolved and recruited—or was recruited by—the language faculty once it appeared fully formed. But such an alternative seems at best a play on definitions, and at worst a disingenuous attempt at safeguarding one's theory (but Chomsky has not argued this—at least not quite: see section 3.6.3). The question is therefore whether Chomsky has convincingly argued that communication is a by-product of the language faculty, or whether the adaptationist account of language-as-communication has been demonstrated.

To bolster his account of communication as a secondary function of language, Chomsky proposes that the non-adaptationist origin of communication is a way to explain the apparent variety of spoken and gestured languages across the world. Yet contrary to what Berwick and Chomsky imply, the variety of externalized languages

is compatible with an adaptationist origin to communication. After all, in an adaptationist account it is the underlying mechanisms that *allow* externalization which are the result of selection pressures, not the externalized and culturally-influenced particular language. Thus there could be a uniform set of mechanisms across the species evolved for externalization of language that would nevertheless provide flexibility with respect to the communication modalities needed to account for variety in externalized language. On the other hand, Chomsky's proposal as well is logically plausible. He argues that the relevant mechanisms either were already present when Merge appeared and therefore any selection pressures applied to them have no relation to the language faculty, or come "for free" with the interface between Merge and the SM system, entailing no pressure from natural selection for communication. The variety of languages thus does not *a priori* tip the balance for or against an account of language origins involving adaptation for communication.

However there is empirical evidence that suggests various aspects of speech production and comprehension did evolve for communication. For instance, research into the FOXP2 gene (Enard, et al., 2002) has shown that this gene is implicated in certain linguistic disabilities, such as "severe articulation disabilities accompanied by linguistic and grammatical impairment", and has "been the target of selection during recent human evolution" (p. 869). The gene therefore is related to externalization of language, and perhaps other aspects of language such as grammar (see Marcus & Fisher, 2003, p. 258 for a summary of the debate surrounding the nature of the neurobiological deficits), as well as having apparently undergone selection pressure, making it an adaptation, presumably for communication. Other aspects as well, such as enhanced breathing control (MacLarnon & Hewitt 1999), and the "Speech is Special" hypothesis (see Pinker & Jackendoff, 2005, pp. 206-208 for a review of the relevant research) have all been advanced to support the hypothesis that the faculty of language evolved through selection pressures related to communication. All the same, Berwick and Chomsky dismiss these claims, spending some time arguing that FOXP2

(at the very least) “does not speak to the question of the core faculty of human language” (2011, p. 34) since it may be related to motor coordination more generally.

Thus

the link between FOXP2 and high-grade serial motor coordination could be regarded as either an opportunistic prerequisite substrate for externalization, no matter what the modality, as is common in evolutionary scenarios, or the result of selection pressure for efficient externalization solutions after Merge arose. (p. 34)

In the first scenario the saltationist story is preserved, since the relevant modifications to FOXP2 are assumed to have happened before Merge and therefore would be part of FLB. However in the second scenario, there is selection pressure related to communication, which happens after the appearance of Merge, apparently invalidating the conclusion that language’s communicative aspects are not an adaptation. Yet Berwick and Chomsky enjoin: “In either case, FOXP2 becomes part of a system extrinsic to core syntax/semantics.” (p. 34) In other words, insofar as they can maintain that the selection pressure for “efficient externalization solutions” did not impact the generative procedure, their saltationist approach is mostly preserved since the major components they believe constitute the language faculty—Merge and the interfaces to the CI and SM systems—remain essentially unchanged.

In sum, the non-adaptationist account of language’s communication aspects has not been demonstrated, and indeed there may be empirical support for the opposite conclusion. If it could be robustly demonstrated that various parts of the internal mechanisms of the language faculty evolved through selection for communication and not any other selection pressure, it would be a hard blow for Chomsky’s hypothesis, eliminating the saltationist aspect. However in many cases the data are inconclusive or at best debatable, as is the case for instance with the FOXP2 research. Piattelli-Palmarini & Uriagereka (2011) argue that many researchers are making inferences regarding the linguistic phenotype of FOXP2 based on coarse brain correlates, with different publications disagreeing over the functions of the implicated

brain regions (p. 100). Furthermore, “without a deeper understanding of development and, especially, of the human-specific aspects of human development, we are not likely to be able to make many definite connections between high-level phenotypes and the role of early-acting regulatory genes [like FOXP2].” (Preuss, 2012, p. 10714) In other words, in many cases more research is needed to understand the link between brain regions, cognitive capacities and linguistic capacity, as well as between genes and brain development, to assess claims regarding the evolutionary development of communication.

3.6.3. No proto-language

Chomsky’s argument regarding the absence of a proto-language depends entirely on his definition of language. For Chomsky, “a person’s language is a computational system of the mind/brain that generates an infinite array of hierarchically structured expressions” (2010, p. 45). Since the generative procedure is thought to be the result of the application of Merge upon lexical items, and Merge could only have appeared in a single step, then by definition language appears with Merge. However, if language is defined instead through aspects relating to communication, then the possibilities for a proto-language are many. Communication through language calls on many properties which potentially could have appeared in various steps, such as displacement, combinatoriality, recursion, reference (although not for Chomsky), symbolism, duality of patterning, etc. Note that even within Chomsky’s framework, at least some of these properties could have succeeded one another, perhaps even before the appearance of Merge, since they do not imply anything with respect to Chomsky’s definition of language. In other words, properties such as displacement or symbolism presumably could have appeared before Merge, since they need not be connected to syntax in any way, relating as they do even to single lexical items.

Chomsky's definition of language thus precludes a proto-language, but does not preclude proto-communication.

If the possibility of a proto-language in Chomsky's sense is quasi logically incoherent, it follows that the "many proposals" invoked which posit a stipulated limitation on Merge are at best implausible. Not surprisingly, no published work seems to actually propose any such limitation on Merge. Behme, in a review of Chomsky (2012), contacted sixteen notable researchers in the field (among others Jackendoff, Newmeyer, Christiansen, Corballis, Lieberman and Bickerton) and asked them if they embraced a theory whereby there were stipulated limitations on Merge at some point in the early stages of its evolution. All the researchers denied their adherence to such a position, with Behme concluding that "the consensus was: "This is a theoretical straw man if I ever saw one"" (Behme, 2013, p. 102). Indeed when proto-language is posited, it is not by a researcher in the MP, and therefore does not typically rely on Merge.

3.7. Chronology revisited

The chronology of language evolution proposed by Chomsky is quite straightforward: first came lexical items in some unknown way, then Merge through a random mutation, then externalization through exaptation of the SM system. Not only does this linear progression seem simplistic compared to the complexities it attempts to capture, it also disregards much of the research in the MP. The greatest lacuna is no doubt the lack of any explanation regarding the appearance of lexical items. In fact, although Chomsky (2010) does advance the non-referentialist thesis regarding words (see sections 2.3.2 and 3.2.5, this paper), and mentions that words are a difficult problem for language evolution (p.57) that must be accounted for (p.62), he does not address the evolution of words any further. On the other hand, Berwick and Chomsky

(2011) do include lexical items in the chronology summarized above. Both articles are problematic for two main reasons, tackled in the following paragraphs: the first is that lexical items and the features they carry play a major role in the generative procedure proposed within the MP and cannot therefore be brushed aside as a detail to be sorted out later. The second is that the role they do play is such that the chronology proposed by Berwick and Chomsky (2011) is, to say the least, unlikely. Regarding this second critique I will first summarize Boeckx's (2011) proposal that it is lexical items rather than Merge that led to the sudden emergence of language. Second, I will follow with Jackendoff's (2010) point that many features are inert without syntax, and to posit their appearance before Merge is therefore at best implausible. Jackendoff nevertheless believes that research within the MP is bound to a saltationist account. Third, I present Clark's (2013) article wherein he argues, against Jackendoff (2010), that the MP is amenable to a gradualist account. Finally, I argue that not only is the MP amenable to a gradualist account, the mechanisms it proposes practically force it into a gradualist scenario regarding the evolution of the language faculty.

3.7.1. The case of the missing lexical items

The role of lexical items, and more specifically lexical features, is major within the MP, and therefore must be addressed in any evolutionary account brought forward. Yet as mentioned previously, the evolution of words is left unexplained in both Chomsky (2010) and Berwick and Chomsky (2011), ostensibly due to a lack of knowledge regarding them. Yet as was shown in the previous chapter, Merge on its own does not account for the generative aspect of language, relying as it does on lexical features to licence further applications of Merge. This problematic absence of any explanation for lexical features in Chomsky's account of language evolution has been pointed out by certain authors. As mentioned in section 2.5.2, Bickerton (2009b)

remarks in passing that the omission of lexical items in fact precludes recursivity itself since “it is the fact that words (and combinations of words) have dependencies that must be filled which drives repeated applications of Merge until the problem is solved—that is, until the complete grammatical sentence is generated.” (footnote 2, pp. 6-7) The recursive process is thus the consequence of the interaction between the combinatorial process and the atoms of that combination, which “force it to result in a hierarchical structure” (footnote 2, p. 7). Jackendoff (2011) puts forward the same argument, simplifying and summarizing it nicely: “recursive structures cannot exist without units to combine” (p. 599). The importance of lexical items could not have been overlooked by Chomsky or Berwick since they each participate actively in research within the MP, both writing articles which rely heavily on features as a mechanism for syntax (cf. Chomsky 1995, 2005; Berwick 2011; Berwick et al. 2011). It is thus curious that such an important aspect of the generative procedure would be left mostly unaccounted for. Needless to say, lexical items and their features need to be included in any account of language evolution within the MP, and their appearance justified through evolutionary biology.

3.7.2. Boeckx: saltationism through lexical items

Of course Berwick and Chomsky (2011) do mention that lexical units need to have evolved, placing their appearance before Merge, a proposal contested by Boeckx (2011). The justification found in Berwick & Chomsky (2011) for the appearance of lexical units is essentially hand waving to parallels between human concepts and concepts possessed by other primates, which presumably suggests that they could have arisen through gradual evolution from those common roots.

Boeckx (2011) argues from within the MP framework, adopting most of the scaffolding used by Chomsky, including the mechanisms posited within the MP, the

idea that externalization of language was secondary and involved no evolution (p.62), as well as the saltationist approach (p.46). Just as Chomsky does, Boeckx proposes that the language faculty evolved in a single step and recruited older cognitive and physiological traits for use in language. However the key component of FLN is not thought to be Merge, but rather the lexical features: “the edge feature [i.e. lexical features], the catalyst for recursive Merge, is the one key property that had to evolve.” (ibid., p. 54) Although Boeckx is very clear that lexical features and Merge are both essential to a fully-formed language faculty, it is unclear when Merge is thought to have evolved. He does argue that “set formation [i.e. the application of Merge] is a very basic computational operation, one that is unlikely to be unique to humans or specific to language” (ibid., p. 52), suggesting that he believes that Merge would have appeared far earlier than lexical features. This relates to a point made in section 2.6, wherein I argued that Merge alone is essentially a simple combinatorial algorithm, likely to be found in other species. It is therefore plausible that Merge is a very old cognitive tool, developed by a common ancestor to many species. Yet despite acknowledging the role of lexical features and downplaying the significance of Merge, Boeckx nevertheless maintains the saltationist position, for much the same reasons Chomsky and Berwick do (adopting the premises of the species-wide language faculty, the human revolution, and communication as a by-product). Alternately, Boeckx’s insistence on the saltationist account despite his divergence from Chomsky’s emphasis on Merge’s appearance may be due to his commitment to the MP, if Jackendoff (2010) is to be believed.

3.7.3. Jackendoff: the MP precludes gradualism

Indeed Jackendoff (2011) argues that “your theory of language evolution depends on your theory of language”, a phrase straightforwardly chosen as the title of the article. According to him, researchers within the MP are committed to a saltationist scenario.

Taking for granted that the MP requires Merge as well as lexical items in order to operate, Jackendoff proposes three scenarios. The first, later proposed by Berwick and Chomsky (2011) is that lexical items came before Merge. Jackendoff (2010) argues that this cannot be the case, aptly remarking that many features are uninterpretable without a role in syntactic constructions. Undeniably certain feature complexes such as “transitive verb” are expressed only within a syntactic construction: “These syntactic aspects of the lexicon are cognitively useful only if there are syntactic trees to insert lexical items into, so it is hard to imagine why or how they should have evolved prior to the advent of syntax.” (ibid., p.70) Berwick and Chomsky’s chronology thus proposes the appearance of features which make no behavioural difference without their interaction with Merge, making their evolution, as well as their spread within the population, highly unlikely. The second proposed scenario is that both Merge and lexical items appeared at once, the lexical features emerging perhaps ‘for free’ through the interfaces to the CI and SM systems. According to Jackendoff this “seems equally hard to imagine” since many features are not properties of either of the interfaces (ibid., p.70). Finally the third scenario is that Merge appeared before the lexical features. Regarding this last proposal, Jackendoff argues that “it’s hard to imagine how syntax [i.e. Merge as a linguistic mechanism] could work without these features, so it looks like they have to join recursion as part of the proposed single step.” (ibid., p.70) In other words, he does not believe that Merge could operate in any significant way without features on which to operate, and therefore a chronology whereby Merge appeared before lexical items is no more likely.

Since Merge requires the lexical items in order to operate as a linguistic mechanism, and the lexical features require Merge in order to be expressed, neither is likely to have appeared and spread within the species before the other. Nevertheless, Jackendoff does offer the caveat that “I’ll be the first to admit that finding something hard to imagine doesn’t make it false” (ibid., p. 70), granting that these scenarios are

not impossible, only very implausible. Consequently, according to Jackendoff, researchers within the MP are practically bound to a saltationist position, and the assumption that both Merge and lexical items appeared in a single step.

3.7.4. Clark: a moderate approach

Contrary to this, Clark (2013) argues that the MP is compatible with a gradualist approach to language evolution. Arguing explicitly against Jackendoff's (2010) article, Clark observes as Jackendoff did that "grammars developed within the minimalist program, even those of the most radical sort, contain at least two components: words (understood as bundles of syntactic features) and the recursive hierarchical operation Merge." (Clark, 2013, p. 188) Consequently, it is possible to posit a multi-step evolution, since more than one mechanism is needed to account for the language faculty. Clark further claims (with apparently no supporting argument contra Berwick & Chomsky, 2011) that externalization as well must be an evolved component, concluding that MP proposals for language evolution "must have involved at least the three steps" (Clark, 2013, p. 190), which is to say the appearance of Merge, lexical items, and externalization. Yet despite the apparent necessity of various steps in Clark's account, his conclusion remains only that the MP can propose a gradualist account and is not bound to a saltationist scenario for language evolution, concluding that "your favoured theory of syntax does not determine your theory of syntactic evolution." (ibid., p. 191)

3.7.5. The MP does not warrant saltationism

Pushing Clark's argument further, I argue that Chomsky (and Berwick) need to abandon the saltationist position in favour of a gradualist account. As was shown in

this chapter, the entire saltationist argument essentially hinges on three premises: (1) that there exists a species-wide language faculty, (2) that there existed a human revolution, explainable only through the appearance of language, and (3) that Merge is the only mechanism which would have needed to evolve. Premise (1) is generally uncontroversial, but implies nothing with respect to the saltationist or gradualist account and can consequently be put aside. Regarding (2), no consensus exists for the human revolution hypothesis at once because the archaeological record is not so clear-cut, and because other factors, namely cultural development, could account for a sudden surge in technological complexity, implying that the impetus for the saltationist explanation is not very strong. But most importantly, premise (3), as pointed out in differing ways by Bickerton (2009b), Boeckx (2011), Jackendoff (2010) and Clark (2013), is simply false, since lexical items and lexical features also need to be accounted for when one takes at face value all the major mechanisms proposed within the MP. Finally, for some researchers saltationist events in evolution are thought to be the exception, and should typically be postulated only when there is significant empirical evidence to warrant them, which is not the case with language. If on the other hand we take for granted that saltations are more common, as certain other researchers do, it still remains to be shown that the trait is so simple as to not allow intermediate steps, without which the assumption that it stems from a small genotype change is likely unwarranted.

3.7.6. An alternative scenario

Considering these points, it is far more reasonable, even from within the MP, to assume that language evolved through successive steps. I propose here a scenario for language evolution which adopts the mechanisms of the MP but shows how they could be used to advance a more plausible, gradualist and adaptationist account for language evolution. Of course this account is highly speculative and is shown here

merely as a way to demonstrate that far from the “simplest” account, Chomsky’s scenario for language evolution disregards many of the interacting mechanisms he himself endorses in the MP, and how taking them into account could lead to a gradualist evolutionary story. My proposed chronology does not include dates and does not specify which parts stem from cultural developments and which parts are biological since these will need to be differentiated through neurobiological, genetic, and linguistic research. Notwithstanding this avowed ignorance, the adaptationist nature of the account implies that it is very reasonable to assume that there could have been many small successive steps to the biological evolution of the language faculty. The objective is thus not to convince the reader that this scenario is the true story of language evolution, but merely that Chomsky need not—and should not—confine himself to a saltationist scenario.

Along the lines proposed by Boeckx (2011), a likely scenario is that Merge appeared very early in evolutionary history, being as it is a very simple combinatorial algorithm, probably shared with other species, suggesting common descent. However without lexical features, it does not produce recursive or hierarchically structured phrases, only combinations of concepts. The concepts themselves may have become combinable through a de-modularisation of the mind, an idea suggested by Mithen (1996), who argues that it is a crucial development in human cognition. Boeckx (2011) as well recruits this idea, claiming that “this ability of building bridges across modules is directly related to language, specifically the ability to lexicalize concepts (uprooting them from their modules) and combine them freely via Merge.” (p. 59) This may also be related to Chomsky’s non-referentialist claim with respect to lexical items, as a de-modularised mind may have gained the ability to apply concepts to other concepts without a necessary external reference relationship. For instance,

“lion” could come to refer not to a specific lion in the world, but instead all the cognitive associations related to experiences of lions.¹⁰

Without features however the combinations of concepts would have been unconstrained, relying only on an interface to the CI system—in other words, the meaning of the concepts—to warrant the pairing. The assumption therefore is that the interface to the CI system was already present since it too has antecedents in other species (Hauser, Chomsky, & Fitch, 2002, p. 1573). No doubt the interface to the SM system as well was available, allowing crude sound-meaning pairs. These interfaces would have permitted feedback on the Merging of concepts since at the CI interface appropriate combinations lead to better planning and understanding, whereas at the SM interface more refined sound-meaning pairings would facilitate communication. This feedback would have led to the ‘crystallisation’ of semantic and phonetic features, transforming the pairings of concepts and their externalization from a more-or-less affair into a precise and determined process. This would make the interfaces and systems more efficient and more precise. For the SM system at least this would allow for co-evolution of the phonetic features and the fine motor control necessary for articulation throughout the entire process of language evolution. This sort of development at the CI interface is more difficult to trace because its definition is far less precise, and knowledge of it is far more indirect. It is nevertheless likely that the same feedback between the CI system and Merge/lexical items would have happened, allowing greater cognitive power to develop.

Once communication of the sort exists, language evolution can be framed in terms of niche construction theory (as Bickerton, 2009a has done): those individuals who

¹⁰ Granted, this is a tenuous argument; but then again, the non-referentialist claim itself is also rather vague and imprecise. Until Chomsky’s position can be clearly articulated and empirical evidence has come to bear, it is difficult to understand quite what is needed to account for its evolution.

cannot communicate at the same level as the others will be less likely to participate effectively in group activities for sustenance (such as hunting and gathering), and could even be socially ostracized from mating. Because of this, there would be strong selection for linguistic and cognitive capacities, leading to increased refinement and complexity of these traits.

Syntactic features would have developed at this point, perhaps through development and refinement of semantic features. Certain syntactic features seem to be loosely related to semantic requirements: for instance the *det* feature mentioned previously could come out of the need to specify the number of objects one is talking about (a semantic requirement), later becoming a syntactic feature necessitating feature-checking. It is thus easy to imagine a transfer from semantic to syntactic features, with the latter proliferating because of the structure and recursion that they allow in language. The labelling algorithm, or projection of the syntactic features during Merge, could have appeared separately as a way of allowing for larger syntactic constructions all the while preserving the structural dependence that comes with syntactic features. Once these larger sentences can be constructed, suppression of duplicates during externalization would become advantageous for the same reasons Chomsky evoked, which is to say the reduction of the computational load associated with fine motor control. With all these mechanisms in place, the language faculty is essentially what it currently is, barring the possibility of additional fine-tuning to the lexical features, interfaces and CI and SM systems.

This is merely a plausible scenario considering the premises studied here, and taking for granted that the mechanisms proposed by the MP are indeed those of the actual language faculty. Note that such an evolutionary history could be compatible with a "human revolution" insofar as niche construction could account for very strong selection forces, driving rapid (albeit gradual) evolution. There remain problems of course, such as what may have been the selection pressures and novelties that pushed

for the initial de-modularisation of the mind, and whether and how the semantic, phonetic and syntactic features all stem from the same genotype, not to mention all the details surrounding the mechanisms within the CI and SM systems. Yet despite these shortcomings, the scenario proposed has the advantage over Chomsky's of including all the mechanisms taken to be necessary for language within the MP, and showing how they need not have appeared all at once, or for completely unknown reasons.

3.8. Conclusion

Chomsky and Berwick's account of the evolution of the language faculty draws from research in paleoanthropology, archaeology, evolutionary biology, and heavily on the findings of the MP. Yet their premises are not always as solid as these authors seem to believe, and the conclusions they draw are more often than not faulty. The most significant problem comes from their own field, and is the omission of an explanation for the evolution of lexical units. Even more problematic is the omission of any mention of lexical features, despite the fact that they are essential to linguistic phenomena and even recursion itself. It is the exclusion of features which allows them to posit the saltationist and non-adaptationist scenario; including them in the evolution of the language faculty radically changes the picture, rendering the saltationist approach highly implausible, and a gradualist scenario far more likely. As such, the saltationist approach advocated by Chomsky and Berwick fails because of the omission of crucial elements from their own research in linguistics: saltationism is therefore not supported by current linguistic theories within the MP.

As previously mentioned, acceptance of the MP is by no means unanimous within linguistics or evolutionary linguistics, and would benefit from a more rigorous account of the evolution of language if it means to convince more researchers of its

validity. In light of this, and as illustrated by the proposed gradualist scenario, if the MP is to contribute in a meaningful way to research in language evolution, a greater emphasis will need to be placed upon elucidating the details of words, concepts, lexical items and lexical features, as well as the relationships between all these, thus covering one of the most egregious lacunas in evolutionary accounts proposed by Chomsky and some of his collaborators.

CONCLUSION

The objective of this research was to describe and critique Chomsky's thoughts regarding language evolution, from their first explicit formulation in Hauser, Chomsky & Fitch (2002), to the latest publications. As was shown, all these articles have significant problems once the concepts and arguments they rely on are carefully unpacked and understood.

HCF (2002) suffers from a few significant setbacks, most notably with respect to recursion. Although put forward as the central property of language and the only component within FLN (along with the interfaces to the systems), recursion remained undefined in the article. Moreover, since recursion is a property and not an algorithm unto itself, claiming that it is the sole property of FLN raises the problem of defining just what algorithm instantiates the recursion. In other words, if the claim was that FLN contains a recursive algorithm, then that algorithm needs to be specified since such algorithms can be vastly different, especially when the definition of recursion remains open for interpretation.

Chomsky subsequently published papers on language evolution in which a specific recursive algorithm is proposed for FLN. Leaning on research within the Minimalist Program, Chomsky (2010; Berwick & Chomsky, 2011) proposes that Merge is the only syntactic operator necessary for inclusion in FLN. Chomsky furthermore takes for granted the existence of a "human revolution" sometime around 75 000 years ago, and that it could only be accounted for through the emergence of language brought about by a biological change. These premises led him to posit a saltationist scenario in which the language faculty appeared in a single step, with the appearance of Merge through a rewiring of the brain, perhaps due to a minor mutation. This language

faculty recruited the pre-existing CI and SM systems, with externalization (communication) appearing later, through no biological evolution.

Yet as was shown, this scenario is problematic for two main reasons. The first is that the human revolution may not be relevant to language evolution, either because the technological progression associated with it may in fact have been gradual—a claim to be sorted out by paleoanthropologists—or because a period of great technological innovation could be explained by other factors, such as cultural development or other biological changes that are not related to language. The second and more significant problem is that the saltationist approach is realistic only insofar as it ignores the significant role played by lexical items and lexical features within the syntactic process. Merge on its own does not account for the formation of all and only well-formed sentences, nor even does it account for recursion itself. By failing to explain how and when lexical items, and more specifically lexical features, have evolved, Chomsky leaves a significant, perhaps even *the* most significant, aspect of the faculty of language out of the picture. With lexical items and features included, the saltationist scenario becomes incredibly unlikely, depending as it does on the idea that the language faculty could hinge on a single language-specific element, namely Merge.

I have proposed an alternative scenario for language evolution that takes into account the bulk of linguistic mechanisms proposed within the MP, but is nevertheless gradual. It is of course highly speculative and should be taken not as claim regarding how the evolution of language truly did happen, but merely as a demonstration that Chomsky need not limit his research into language evolution to a saltationist approach. This research thus opens up possibilities for investigations into gradualist accounts of language evolution for the MP. It also demonstrates that, should researchers within the MP want to contribute meaningfully to an evolutionary account

of language evolution, more research will be needed regarding the origin and role of lexical features.

Finally, this research has not touched upon many interesting and related issues which could have a significant impact on the MP and its involvement in language evolution debates. For instance, the emphasis on optimality within the faculty of language has peculiar consequences when applied to language evolution: in Chomsky's approach, it was the driving force behind the justification of a saltationist scenario. But is this optimality truly a good way of approaching reconciliation with biology? Boeckx (2011) points out that "it is emphatically not the sort of optimization that ultra-Darwinists like Dawkins advocate" (p. 56). It is instead the sort that is thought to stem from principles of physics or chemistry ("laws of nature" as Berwick & Chomsky put it: 2011, p. 30)), not biology. Among others, "principles of structural architecture and developmental constraints" (Chomsky, 2005, p. 6) are thought to fashion the language faculty, making its optimality something more akin to "a snowflake" (Chomsky, 2010; p. 59) than a biological organ. Researchers within the MP (Chomsky, 2010; Boeckx, 2011; Berwick & Chomsky, 2011) often call upon findings in evolutionary developmental (evo-devo) biology to justify such claims but it is not clear that the parallels are justified (c.f. Benítez-Burraco & Longa, 2010). Needless to say, these issues relate directly to language evolution and the MP, and warrant further investigation.

BIBLIOGRAPHY

- Behme, C. (2013). Review of *The Science of Language* (Chomsky, 2012). *Philosophy in Review*, XXXIII(2), 100-103.
- Behme, C. (2014, July). *A Potpourri of Chomskyan Science*. Retrieved August 27, 2014, from LingBuzz: <http://ling.auf.net/lingbuzz/001592/>
- Benítez-Burraco, A., & Longa, V. M. (2010). Evo-Devo — Of Course, But Which One? Some Comments on Chomsky's Analogies between the Biolinguistic Approach and Evo-Devo. *Biolinguistics*, 4(4), 308-323.
- Berwick, R. C. (2011). Syntax Facit Saltum Redux: Biolinguistics and the Leap to Syntax. In A. Di Sciullo, & C. Boeckx (Eds.), *The Biolinguistic Enterprise* (pp. 65-99). New York: Oxford University Press.
- Berwick, R. C., & Chomsky, N. (2011). The Biolinguistic Program: The Current State of Its Development. In A. Di Sciullo, & C. Boeckx (Eds.), *The Biolinguistic Enterprise* (pp. 19-41). New York: Oxford University Press.
- Bickerton, D. (2009a). *Adam's Tongue: How Humans Made Language, How Language Made Humans*. New York: Hill and Wang.
- Bickerton, D. (2009b). Recursion: Core of Complexity or Artifact of Analysis? In T. Givón, & M. Shibatani (Eds.), *Syntactic Complexity: Diachrony, Acquisition, Neuro-cognition, Evolution* (pp. 531-544). Philadelphia: John Benjamins Publishing Company.
- Bickerton, D. (2009c). Syntax for Non-syntacticians. In D. Bickerton, & E. Szathmáry (Eds.), *Biological Foundations and Origin of Syntax* (pp. 3-13). Cambridge, Massachusetts: MIT Press.
- Boeckx, C. (2006). *Linguistic Minimalism: Origins, Concepts, Methods, and Aims*. New York: Oxford University Press.
- Boeckx, C. (2011). Some Reflections on Darwin's Problem in the Context of Cartesian Biolinguistics. In A. Di Sciullo, & C. Boeckx (Eds.), *The Biolinguistic Enterprise* (pp. 42-64). New York: Oxford University Press.

- Botha, R. P. (1998). Slaying 'the Nonselectionist Theory' of Language Evolution. *Stellenbosch Papers in Linguistics*, 31, 26-60.
- Chomsky, N. (1957). *Syntactic Structures* (11th ed.). Paris: Mouton.
- Chomsky, N. (1971). *Problems of Knowledge and Freedom*. London: Fontana.
- Chomsky, N. (1988). *Language and the Problems of Knowledge*. Cambridge, MA: MIT Press.
- Chomsky, N. (2000a). *New Horizons in the Study of Language and Mind*. New York: Cambridge University Press.
- Chomsky, N. (2000b). *The Architecture of Language*. New York: Oxford University Press.
- Chomsky, N. (2005, Winter). Three Factors in Language Design. *Linguistic Inquiry*, 36(1), 1-22.
- Chomsky, N. (2010). Some Simple Evo Devo Theses: How True Might They Be for Language? In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language* (pp. 45-62). New York: Cambridge University Press.
- Chomsky, N. (2011, April 8). Language and the Cognitive Revolution(s). *Andrew Brook Distinguished Lecture Series* (<https://youtu.be/XbjVMqOk3uc>). Carleton University.
- Chomsky, N. (2012). *The Science of Language*. Cambridge: Cambridge University Press.
- Chomsky, N. (2013, October 25). *What is language, and why does it matter?* (UQAM, Producer) Retrieved 08 27, 2014, from UQAM.tv: <http://youtu.be/K-mWHbpT6Lc>
- Chouard, T. (2010). Evolution: Revenge of the Hopeful Monster. *Nature*, 463, 864-867.
- Churchland, P. S. (1986). *Neurophilosophy: Toward a Unified Science of the Mind/Brain*. Cambridge, MA: MIT Press.
- Clark, B. (2013). Syntactic Theory and the Evolution of Syntax. *Biolinguistics*, 7, 169-197.
- Cohen, H. (2013). Historical, Darwinian, and Current Perspectives on the Origin(s) of Language. In C. Lefebvre, B. Comrie, & H. Cohen (Eds.), *Studies in Language*

Companion: New Perspectives on the Origins of Language (Vol. 144, pp. 3-30). Philadelphia: John Benjamins.

Collins, C., & Stabler, E. P. (2011). *A Formalization of Minimalist Syntax*. Retrieved August 2014, from Lingbuzz: <http://ling.auf.net/lingbuzz/001691>

Corballis, M. C. (2011). *The Recursive Mind: The Origins of Language, Thought, and Civilization*. Princeton: Princeton University Press.

Cosmides, L., & Tooby, J. (1987). From Evolution to Behavior: Evolutionary Psychology as the Missing Link. In J. Dupré (Ed.), *The Latest on the Best. Essays on Evolution and Optimality* (pp. 277-306). Cambridge MA: MIT Press.

Cosmides, L., & Tooby, J. (1997). *Evolutionary Psychology Primer*. Retrieved October 20, 2013, from Center for Evolutionary Psychology: <http://www.cep.ucsb.edu/primer.html>

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge, MA: Cambridge University Press.

Darwin, C. (1859). *On the Origin of Species* (1st ed.). London: John Murray.

Di Scullio, A., & Boeckx, C. (2011). Introduction: Countours of the Biolinguistic Agenda. In A. Di Scullio, & C. Boeckx (Eds.), *The Biolinguistic Enterprise* (pp. 1-16). New York: Oxford University Press.

Dunbar, R. (1996). *Grooming, Gossip, and the Evolution of Language*. Cambridge, MA: Harvard University Press.

Durrant, R., & Haig, B. D. (2001). How to Pursue the Adaptationist Program in Psychology. *Philosophical Psychology*, 14(4), 357-380.

Enard, W., Przeworski, M., Fisher, S. E., Lai, C. S., Wiebe, V., Kitano, T., et al. (2002, August 22). Molecular Evolution of FOXP2, a Gene Involved in Speech and Language. *Nature*, 418, 869-872.

Everett, D. L. (2005, August-October). Cultural Constraints on Grammar and Cognition in Pirahã. *Current Anthropology*, 46(4), 621-646.

- Everett, D. L. (2006, July). Biology and Language: Response to Anderson & Lightfoot. *Journal of Linguistics*, 42(2), 385-393.
- Everett, D. L. (2007). *Cultural Constraints on Grammar in Pirahã: A Reply to Nevins, Pesetsky, and Rodrigues*. Retrieved May 25, 2014, from LingBuzz: <http://ling.auf.net/lingbuzz/000427>
- Everett, D. L. (2007, June 11). *Recursion and Human Thought*. Retrieved May 28, 2014, from Edge.org: <http://edge.org/conversation/recursion-and-human-thought>
- Everett, D. L. (2009, December). Letters to Language. *Language*, 85(4), 753.
- Everett, D. L. (2009). Pirahã Culture and Grammar: A Response to Some Criticisms. *Language*, 85(2), 405-442.
- Fitch, W. T. (2010). Three Meanings of "Recursion": Key Distinctions for Biolinguistics. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language* (pp. 73-90). Cambridge: Cambridge University Press.
- Fitch, W. T., Hauser, M. D., & Chomsky, N. (2005). The Evolution of the Language Faculty: Clarifications and Implications. *Cognition*, 97, 179-210.
- Gould, S. J. (1980, June/July). The Return of Hopeful Monsters. *Natural History*, 86, 22-30.
- Gould, S. J. (2002). Punctuated Equilibrium's Threefold History. In *The Structure of Evolutionary Theory* (pp. 1006-1021). Cambridge, MA: Harvard University Press.
- Gould, S. J., & Eldredge, N. (1977). Punctuated Equilibria: the Tempo and Mode of Evolution Reconsidered. *Paleobiology*, 3(2), 115-151.
- Gould, S. J., & Lewontin, R. C. (1979). The Spandrels of San Marco and the Panglossian Paradigm: A Critique of the Adaptationist Programme. *Proceedings of the Royal Society of London, Series B*, 205(1161), 581-598.
- Gould, S. J., & Vrba, E. S. (1982). Exaptation - A Missing Term in the Science of Form. *Paleobiology*, 8(1), 4-15.
- Hauser, M. D. (2010). On Obfuscation, Obscurantism, and Opacity: Evolving Conceptions of the Faculty of Language. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language* (pp. 91-99). Cambridge: Cambridge University Press.

- Hauser, M. D., Chomsky, N., & Fitch, W. (2002, November 22). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science*, 298, 1569-1579.
- Hoffecker, J. F. (2007, December). Representation and Recursion in the Archeological Record. *Journal of Archeological Method and Theory*, 14(4), 359-387.
- Holmberg, A. (2000). Am I Unscientific? A Reply to Lappin, Levine and Johnson. *Natural Language & Linguistic Theory*, 20(4), 837-842.
- Huxley, T. H. (1859, November 23). *Letter from Huxley, T. H. to Darwin, C. R.* Retrieved 05 25, 2014, from Darwin Correspondence Project: <http://www.darwinproject.ac.uk/entry-2544>
- Jackendoff, R. (2002). *Foundations of Language*. New York: Oxford University Press.
- Jackendoff, R. (2010). Your Theory of Language Evolution Depends on Your Theory of Language. In R. K. Larson, V. Déprez, & H. Yamakido (Eds.), *The Evolution of Human Language* (pp. 63-72). Cambridge: Cambridge University Press.
- Jackendoff, R. (2011, September). What is the Human Language Faculty? Two Views. *Language*, 87(3), 586-624.
- Jackendoff, R., & Pinker, S. (2005). The Nature of the Language Faculty and its Implications for the Evolution of Language (Reply to Fitch, Hauser and Chomsky). *Cognition*, 97, 211-225.
- Karlsson, F. (2010). Syntactic Recursion and Iteration. In H. van der Hulst (Ed.), *Recursion and Human Language* (pp. 43-67). New York: Mouton de Gruyter.
- Kenneally, C. (2007). *The First Word: The Search for the Origins of Language*. New York: Penguin Books.
- Kinsella, A. R., & Marcus, G. F. (2009). Evolution, Perfection, and Theories of Language. *Biolinguistics*, 186-212.
- Laland, K. N., & Brown, G. R. (2011). *Sense & Nonsense: evolutionary perspectives on human behaviour*. New York: Oxford University Press.
- Lappin, S., Levine, R. D., & Johnson, D. E. (2000). The Structure of Unscientific Revolutions. *Natural Language & Linguistic Theory*, 18(3), 665-671.

- Leroi-Gourhan, A. (1964). *Le Geste et la Parole*. Paris: Albin Michel.
- Lewontin, R. C. (1970). The Units of Selection. *Annual Review of Ecology and Systematics*, 1, 1-18.
- Lewontin, R. C. (1998). The Evolution of Cognition: Questions We Will Never Answer. In *An invitation to cognitive science* (Vol. 4, pp. 107-131). Cambridge, MA: MIT Press.
- Liberman, P. (2002). On the Nature and Evolution of the Neural Bases of Human Language. *Yearbook of Physical Anthropology*, 45, 36-62.
- Lieberman, P., Klatt, W. L., & Wilson, W. A. (1969). Vocal Tract Limitations on the Vowel Repertoires of Rhesus Monkey and Other Nonhuman Primates. *Science*, 164, 1185-1187.
- Machery, E. (2005). Concepts Are Not a Natural Kind. *Philosophy of Science*, 72, 444-467.
- Marcus, G., & Fisher, S. (2003). FOXP2 in Focus: What Can Genes Tell Us About Speech and Language? *Trends in Cognitive Sciences*, 257-262.
- Mellars, P. A., & Stringer, C. (Eds.). (1989). *The Human Revolution. Behavioural and Biological Perspectives in the Origins of Modern Humans*. Edinburgh: Edinburgh University Press.
- Meunier, J.-G. (2006). Pensée et Langage Dans la Philosophie Herméneutique Frégéenne. In F. B. Illa (Ed.), *A Proposito di Rappresentazioni. Alla Ricerca del Senso Perduto* (pp. 25-84). Bologna: Pendragon.
- Millikan, R. (1984). *Language, Thought, and Other Biological Categories*. Cambridge, MA: MIT Press.
- Mithen, S. J. (1996). *The Prehistory of the Mind: the Cognitive Origins of Art, Religion and Science*. London: Thames and Hudson.
- Narita, H., & Fujita, K. (2010). A Naturalist Reconstruction of Minimalist and Evolutionary Biolinguistics. *Biolinguistics*, 356-376.
- Nevins, A., Pesetsky, D., & Rodrigues, C. (2009, June). Pirahã Exceptionality: A Reassessment. *Language*, 85(2), 355-404.

- Ostrom, J. (1979). Bird flight: How Did it Begin? *American Scientist*, 67, 46-56.
- Parker, A. R. (2006). Evolving the Narrow Language Faculty: was Recursion the Pivotal Step? *Proceedings of the 6th International Conference on the Evolution of Language*, (pp. 239-246). Illinois.
- Parker, G. A., & Maynard Smith, J. (1990). Optimality Theory in Evolutionary Biology. *Nature*, 348, 27-33.
- Piattelli-Palmarini, M. (1980). *Language and Learning: The Debate Between Jean Piaget and Noam Chomsky*. Cambridge: Harvard University Press.
- Piattelli-Palmarini, M. (1989, February). Evolution, Selection and Cognition: From "Learning" to Parameter Setting in Biology and in the Study of Language. *Cognition*, 31(1), 1-44.
- Piattelli-Palmarini, M., & Uriagereka, J. (2011). A Geneticist's Dream, a Linguist's Nightmare: The Case of FOXP2. In *The Biolinguistic Enterprise* (pp. 100-125). New York: Oxford University Press.
- Pinker, S., & Bloom, P. (1990). Natural Language and Natural Selection. *Behavioral and Brain Sciences*, 13(4), 707-784.
- Pinker, S., & Jackendoff, R. (2005). The Faculty of Language: What's Special About It? *Cognition*, 95, 201-236.
- Postal, P. M. (2004). *Skeptical Linguistics Essays*. New York: Oxford University Press.
- Preuss, T. M. (2012, June 26). Human Brain Evolution: From Gene Discovery to Phenotype Discovery. *PNAS*, 109, suppl. 1, 10709-10716.
- Rizzi, L. (2009). Some Elements of Syntactic Computations. In D. Bickerton, & E. Szathmáry (Eds.), *Biological Foundations and Origin of Syntax* (pp. 63-87). Cambridge, Massachusetts: MIT Press.
- Savage-Rumbaugh, S. E., & Lewin, R. (1994). *Kanzi: The Ape at the Brink of the Human Mind*. New York: Wiley.
- Savage-Rumbaugh, S. E., Shanker, S., & Taylor, T. J. (1996). Apes with Language. *Critical Quarterly*, 45-57.

- Savage-Rumbaugh, S. E., Shanker, S., & Taylor, T. J. (1998). *Apes, Language, and the Human Mind*. New York: Oxford University Press.
- Shapiro, J. A. (2011). *Evolution: A View from the 21st Century*. Upper Saddle River, New Jersey: FT Press Science.
- Sloan, P. (2010, Fall). *Evolution*. (E. N. Zalta, Editor) Retrieved from The Stanford Encyclopedia of Philosophy:
<http://plato.stanford.edu/archives/fall2010/entries/evolution/>
- Stearns, S. C., & Hoekstra, R. F. (2000). *Evolution: an Introduction*. New York: Oxford University Press.
- Steels, L. (2000, September). Language as a Complex Adaptive System. (D. K. Schoenauer, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, & H.-P. Schwefel, Eds.) *Proceedings of the 6th International Conference on Parallel Problem Solving from Nature*, 1917, 17-26.
- Tallerman, M., Newmeyer, F., Bickerton, D., Bouchard, D., Kaan, E., & Rizzi, L. (2009). What Kinds of Syntactic Phenomena Must Biologists, Neurobiologists, and Computer Scientists Try to Explain and Replicate? In D. Bickerton, & Eörs Szathmáry (Eds.), *Biological Foundations and Origin of Syntax* (pp. 135-157). Cambridge, MA: The MIT Press.
- Terrace, H. S., Petitto, L. A., Sanders, R. J., & Bever, T. G. (1979). Can an Ape Create a Sentence? *Science*, 206, 891-902.
- Theissen, G. (2009). Saltational Evolution: Hopeful Monsters are Here to Stay. *Theory in Biosciences*, 128, 43-51.
- Tinbergen, N. (1963). On the Aims and Methods of Ethology. *Zeitschrift für Tierpsychologie*, 20, 410-433.
- Viger, C. (2007). The Acquired Language of Thought Hypothesis: A Theory of Symbol Grounding. *Interaction Studies*, 18, 125-142.
- Williams, G. C. (1966). *Adaptation and Natural Selection* (8th ed.). Princeton, NJ: Princeton University Press.
- Wimpenny, J. H., Weir, A. A., Clayton, L., Rutz, C., & Kacelnik, A. (2009). Cognitive Processes Associated with Sequential Tool Use in New Caledonian Crows. *PLoS ONE*, 4(8).