Cartesian Linguistics: From Historical Antecedents to Computational Modeling

by

Christina Behme

Submitted in partial fulfilment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
June 2011

DALHOUSIE UNIVERSITY

DEPARTMENT OF PHILOSOPHY

The undersigned hereby certify that they have read and recommend to the Faculty of

Graduate Studies for acceptance a thesis entitled "*Cartesian Linguistics: From Historical

Antecedents to Computational Modeling*" by Christina Behme in partial fulfillment of the

requirements for the degree of Doctor of Philosophy.

Dated: June 07, 2011

External Examiner: _____

Research Supervisor: _____

Examining Committee: _____

_____

_____

Departmental Representative: _____

DALHOUSIE UNIVERSITY

DATE: June 7, 2011

AUTHOR: Christina Behme

TITLE: *Cartesian Linguistics: From Historical Antecedents to Computational Modeling*

DEPARTMENT OR SCHOOL: Department of Philosophy

DEGREE: PhD   CONVOCATION: October   YEAR: 2011

_____
Signature of Author

Table of Contents

Abstract

Chomsky's Cartesian Linguistics frames into his linguistic work and the resulting debates between rationalists and empiricists. I focus on the following key aspects: (i) the historic connection of Cartesian Linguistics to previous linguistic theorizing, (ii) the development of Chomsky's own theorizing, (iii) the empirical work addressing the problem of language acquisition and (iv) the problem of computational modeling of language learning. Chomsky claims that his view is situated within a rationalist Cartesian tradition and that only rationalists will be able to account fully for all aspects of human language. My thesis challenges both claims.

I found only remote connections between Cartesian and Chomskyan commitments. Chomsky holds that (i) language is species-specific, (ii) language is domain-specific, and (iii) language acquisition depends on innate knowledge. Descartes accepted (i), but argued that language is an indicator of domain-general intelligence. Innate resources play a different role for language acquisition for Chomsky and for Descartes.

Chomsky revived linguistics during the 1950s by promising to make it a rigorous part of the biological sciences. However, his work has not resulted in a better understanding of language acquisition and use. Key concepts like 'innateness', 'Universal Grammar' and 'Language Acquisition Device' remain in need of precise definition, and the Poverty of the Stimulus Argument does not rule out data-driven domain-general language acquisition.

Empirical work in developmental psychology has demonstrated that children acquire and practice many language-related cognitive abilities long before they produce their first words. Chomsky's dictum that language learning is uniform across the species and invariably follows genetically determined stages remains empirically unconfirmed. Computational modeling has accounted for some internal structure of language acquisition mechanisms and simulates the specific conditions under which children learn language. Contemporary models use samples of child-directed-speech as input and have replicated numerous aspects of human performance.

Given my findings I suggest that Chomskyan linguistics is not Cartesian in substance or in spirit. Descartes was wary of those "who take no account of experience and think that truth will spring from their brains like Minerva from the head of Jupiter" (CSM I, p. 21). His science relied on sense experience (empiricism) and deduction (rationalism) and a truly Cartesian Linguistics will revive this part of the Cartesian tradition.

.

Chapter 1. Introduction

Cartesian Linguistics, originally published with the purpose of deepening "our understanding of the nature of language and the mental processes and structures that underlies its use and acquisition" (Chomsky, 1966, p. ix), has generated controversy from the time it was first released in 1966 to its recent 3rd edition in 2009. On the one hand we find exuberant enthusiasm, calling it "an intellectual tour de force... an unprecedented and - so far - unequalled linguistic-philosophical study of linguistic creativity and the nature of the mind" (McGilvray, 2009, p. 1). On the other hand it has been severely criticized "Chomsky's version of the history of linguistics… is fundamentally false from beginning to end - because the scholarship is poor, because the texts have not been read, because the arguments have not been understood…" (Aarsleff, 1971, p.584). In my thesis I will evaluate the arguments of both sides.

I will use *Cartesian Linguistics* as a framework for an inquiry into the linguistic work of Noam Chomsky. Chomsky's work has been tremendously influential, and a comprehensive evaluation of it would exceed the scope of any dissertation. I will focus on several key aspects here: (i) the historic connection of *Cartesian Linguistics* to previous linguistic theorizing, (ii) the development of Chomsky's own theorizing, (iii) the empirical work addressing the problem of language acquisition, and (iv) the problem of computational modeling of language learning. Each of these themes will be dealt with in one chapter. Before delving into

the substantive part of my thesis a few preliminary remarks are in order.

Essentially, the philosophical positions on language acquisition taken by Chomsky and his critics fall into two broad groups. Either all our linguistic knowledge comes directly (perception) or indirectly (inference, induction) from sense experience, or at least some of our linguistic knowledge is innate. In philosophy proponents of the first view are empiricists, and rationalists hold the second view. Similarly, in psychology, empiricism (sometimes called the 'blank slate' or *tabula rasa* view) holds that virtually everything is learned through interaction with the environment, and nativism is the view that certain skills or abilities are hard wired into the brain at birth. Currently no one holds either the pure empiricist or pure rationalist view, and I will introduce more nuanced positions as I discuss proponents of either tradition.

Chomsky has defended throughout his career a rationalist/nativist view of language acquisition and language use. He has claimed that this view can be traced back to important linguistic, philosophical, and scientific precursors. According to Chomsky these have been largely neglected:

> Modern linguistics, however, has self-consciously dissociated itself from traditional linguistic theory and has attempted to construct a theory of language in an entirely new and independent way. The contributions to linguistic theory of an earlier European tradition have in general been of little interest to professional linguists, who have occupied themselves with quite different topics within an intellectual framework that is not receptive to the problems that gave rise to earlier linguistic study or the insights that it achieved; and these contributions are by now largely unknown or regarded with unconcealed contempt. (Chomsky, 1966, p. 1)

These remarks were written in 1966 but are repeated, without comment, in 2002 and 2009. I will show that while they may have had some justification in the 1960s, they are not applicable to current linguistics. When dealing with the historic antecedents I will focus on two distinct but closely interrelated aspects. First I will give a brief account of Chomsky's original proposals as expressed in the 1966 edition of *Cartesian Linguistics*. Second I will focus on how the term 'Cartesian Linguistics' has been used by Chomsky and some of his close followers over the last five decades to argue for the superiority of their linguistic theorizing.

Chomsky (1966) claim that some of the allegedly neglected insights have been those of René Descartes. I will give a detailed account of the role 'innate ideas' and 'poverty of the stimulus' arguments (both prominent elements in support of contemporary nativist/rationalist positions) played in Descartes' writings. It will become evident that insofar as Descartes concerned himself specifically with linguistic theory, his insights are equally compatible with the theories of contemporary nativists and empiricists. Furthermore, a close reading of Descartes' own work makes it dubious that Chomsky's work can be traced back to a coherent rationalist tradition of which Descartes was one important founder.

When discussing Descartes' writings, my emphasis will be on issues that are specifically important to Chomsky's linguistic theories. Chomsky holds that Descartes' commitment to innate ideas foreshadows his own postulation of an innate domain specific language acquisition device (LAD). Focusing on three fundamental claims of Chomsky that (i) language is species specific, (ii) language is domain-specific, and (iii) language depends on innate knowledge, I will inquire

whether or not Descartes was committed to the view Chomsky attributes to him. I will provide textual evidence supporting the conclusion that, while Descartes believed that language is species specific, he was committed neither to the view that language is domain specific nor to the view that language acquisition depends on innate knowledge in a sense that is compatible with Chomsky's use of these terms. Further, I will show that Chomsky misunderstands important arguments for Descartes' belief that animals could not acquire a language of the human type.

Descartes' writings demonstrate that he believed that language is species specific. Virtually all humans acquire language and use it regardless of a wide range of differences in their age, health, and intelligence. On the other hand, no animal has cognitive abilities that would allow the use of language. The fact that language is species specific could be explained in two different ways. The first scenario is that only humans have a domain specific language faculty. In this case it would be possible that an animal that had implanted an artificial language faculty would behave in ways that are indistinguishable from a human being. In the second scenario language is an indicator of general intelligence (thought and reason in Descartes' terminology). In this case it would not be possible to 'construct' an artificial language faculty that is independent of 'general intelligence'. I will provide evidence for my claim that Descartes was committed to the second scenario.

When Descartes discusses the differences between animals and humans he stresses repeatedly that very little reason is needed to use language. This could indicate he believed, like Chomsky, that humans have a domain specific language

faculty that is independent of 'general intelligence'. I will show that the purpose of Descartes' comparative examples is to show that the most fundamental difference between humans and animals is that only humans possess a mind. Language requires a rational mind, and, according to Descartes, animals lack such a mind (Gunderson, 1964; Miel, 1969). Having a rational mind does not entail having domain-specific mental faculties. In fact it is well established that for Descartes minds are indivisible: "we cannot understand a mind except as being indivisible… we cannot conceive of half a mind" (CSM II, p. 9) and that the essence of mind is thought. I will show that these strong commitments prevent Descartes from holding that language is domain-specific. Seemingly, Chomsky does not understand this commitment of Descartes (e.g., Chomsky, 1975b, 2010b). For Chomsky, animals have some form of 'general intelligence'. But they do not acquire language because they lack a domain-specific language faculty. Mistakenly, he believes Descartes shares this view. However, Descartes believes that animals do not acquire language because they do not have any intelligence. Thus, for Descartes language is an essential characteristic of humans (species-specific), but it is only a domain-general indicator of a rational mind.

Next, I will show that the textual evidence supports a possibly surprising view about Descartes' commitments regarding language acquisition. He states that language "can be acquired without any process of reasoning... [based] on experience alone" (CSM, II, p. 403), that we learn language by connecting words with their meanings and remembering later upon encountering words which things they signify and vice versa (CSMK, III, p. 307) at a time when our thoughts are

'confused' and based of 'misconceptions'. Of course, the language we acquire under such circumstances is not a perfect tool for the correct expression of our thoughts. But while engagement with philosophical or scientific work requires that we employ new ways of thinking Descartes does not suggest that our language needs to be changed. Descartes also does not hold that language acquisition is a mechanical process of brain maturation in accordance with 'deterministic physical principles', as Chomsky (2010b) incorrectly suggests.

After establishing these points about language acquisition I will show that my reading of Descartes is compatible with his theory of innate ideas. Discussing the frequently cited passages in *Comments on a Certain Broadsheet* (CSM I, p. 304), where Descartes asserts that all ideas are innate, I will show that a careful reading of the context reveals that Descartes' main goal here was to refute a scholastic account of perception. On this interpretation what is innate is not the content of sensory ideas but the faculty of sense perception. I will show how this interpretation allows for a coherent Cartesian account of language acquisition. On my reading the role that Cartesian innate ideas play for language acquisition is very different from the role innate knowledge plays for Chomskyan accounts. For these reasons I suggest that, from a perspective of 'the history of ideas', it is quite misleading to call Chomsky's approach to linguistics Cartesian. I will discuss some of the reasons Chomsky provides for interpreting Descartes in a very unconventional way and suggest that the result is an uneasy hybrid view that does not reflect genuine Cartesian commitments.

The next step will be to connect the historic antecedents of *Cartesian*

Linguistics to contemporary debates in linguistics, philosophy of language and developmental psychology. The linguistic debates have changed considerably since the 17th century and even since the time of the first edition of *Cartesian Linguistics*. Today virtually all researchers agree that extreme rationalism/nativism (all knowledge is innate) is as implausible as extreme empiricism (nothing is innate). Thus, recently the debate has shifted and concerns mainly how much of our knowledge extends beyond our sense experience. Empiricist and rationalist researchers inquire about the character of the interaction with our linguistic environment and the nature of the mechanisms that allow us to acquire linguistic knowledge. As we learn more about the structure of the human brain and the learning mechanisms available to children, we can develop more tools to (i) clarify boundaries between innate and acquired linguistic knowledge and (ii) evaluate how pre-linguistic infants can extract information from their environment. Furthermore, research in linguistics has provided new insights regarding the status of core issues such as types of grammars, recursion and the role of linguistic intuitions and empirical testing. Recently, computational models of language acquisition have provided additional means for testing different hypotheses. The information gained from these different sources has important implications for subsequent philosophical theorizing.

In chapter 3 I will give an overview of the evolution of Chomsky's theorizing during the past six decades. My inquiry is guided by McGilvray's (2009) depiction of steady progress and overwhelming success and focuses on the work of Chomsky and his closest followers. This means I will use the term 'Chomskyan'

only to refer to this restricted group, not to the much larger community of linguists that have been influenced by Chomsky in some way but departed in important points from the views he defends. I chose this focus because Chomsky has been such an influential figure and is widely considered as instrumental to the cognitive revolution of the 1950s even by those who disagree with his views (e.g., Sampson, 1980; Katz, 1981, 1996; Seuren, 1998; Boden 2006). Linguists, cognitive scientists and philosophers continue to look to him for inspiration. I will provide comprehensive evidence for my conclusion that this trust in Chomsky's intellectual leadership is no longer justified.

The focus will be on McGilvray's claim that Chomsky's work has situated linguistics firmly within the natural sciences and provided a better understanding of language acquisition and language use. I will discuss how the core claims of Chomskyan theorizing have changed over time and show how this affected his attitude towards 'traditional scientific practice'. I will then introduce several criticisms of Chomsky's work, concerning methodological and conceptual issues and show that Chomsky has not been able to address these criticisms satisfactorily. Further, I will evaluate the contributions that Chomsky's work made to our understanding of how children learn language. Chomsky's definition of what language is, and thus what children acquire when they learn language, has changed considerably. I will track these changes and argue that they have not led to a better understanding of language acquisition.

According to McGilvray, Chomsky's theorizing has made steady progress, and his theories allow us to account for 'everyday linguistic creativity', which is

acquired by children at an early age (four years according to McGilvray, 2009, p. 7). He claims that Chomsky's work has demonstrated that innate concepts alone can explain the uniform acquisition of language across human populations in spite of alleged poverty of stimulus facts. On this account the child does not learn language but accesses what is innately available to her: "...the mind's concepts and the way of putting them together in language and thought are largely innate" (Ibid., p.6), "the only way to explain the early appearance of creativity is to assume innateness of both concepts and combinatorial principles" (Ibid., p. 7). "Innateness provides a basis for understanding one another even at a young age" (Ibid. p.7), and again "concepts and language are somehow implicit in some kind of natural 'mechanism' of the human body-mind, under (partial) control of the genome and the course of development it controls" (Ibid., pp. 18 - 19). McGilvray asserts that Chomsky's theory is simple, objective, descriptively and explanatorily adequate. It accommodates the science of language to biology and makes steady progress. My analysis will show that McGilvray's account exaggerates the success of Chomsky's work and fails to acknowledge shortcomings and failures.

To show why I disagree with McGilvray's evaluation I will give a detailed account of the evolution of Chomsky's linguistic theories. To situate Chomsky's claims about language acquisition we need to understand what his theories commit him to. Fiona Cowie (1999) has argued that Chomsky's nativism is a complex view that commits him to several (logically independent) claims. First, Chomsky endorses a form of representationalism. He holds that contentful mental states (representations) play a crucial role in the production and explanation of linguistic

behaviour. Human minds contain a finite stock of simple ideas and rules for combining them into more complex thoughts. Second, Chomsky believes that biological bondedness places restrictions on the kinds of contents our thoughts can have. Thus, our biology constrains the possible hypotheses that language learners can entertain. Third, Chomsky holds that domain specificity limits the hypotheses children can entertain during language learning even further by principles that are specific to the linguistic domain. Fourth, Chomsky holds that innateness of language implies that the constraints on language learning are innately provided. He has suggested that these constraints have to be encoded by our genes. Fifth, and probably most importantly, for Chomsky Universal Grammar (UG) identifies the nature of the domain specific language faculty. "Universal Grammar is a theory about those features that natural languages have qua natural languages. It describes the class of 'linguistic universals', the properties that are common to all languages" (Cowie, 1999, p. 156).

Early in his linguistic career Chomsky focused on syntax and grammar (e.g., Chomsky 1951, 1956, 1957, 1959, 1965a,b, 1966, 1968). Some important contributions of Chomsky's early work were the proposals (i) that human languages have syntactic universals, (ii) a grammar defines the class of grammatical sentences, and (iii) the universals define a range of humanly-possible grammars (and by implication rule out any logically-possible grammar not contained within that range). Chomsky claimed that human languages cannot be generated by simple constituency grammars alone and proposed that an additional series of transformational rules is needed to generate all grammatical sentences of human

languages. This early work contributed to clarifying important conceptual issues, provided a scientific framework for linguistics and had an impact that reached far beyond linguistics. I will discuss some of these early contributions and show how they were relevant to his theories of language acquisition.

During the following decades Chomsky's "has overturned and replaced his own established systems with startling frequency" (Smith, 1999, p. 1). I will give an overview of some of the milestones of these conceptual changes. Following Lenneberg (1967) Chomsky proposed that the true subject of linguistic inquiry should be the "deep structure" of language (Chomsky, 1970, 1973). He suggested that this deep structure is the same for all human languages and only indirectly reflected in the 'surface structures' of languages such as English. Chomsky suggested that a set of transformational rules converts deep structures to surface structures. Over the next two decades the complexity of these rules increased continually. Other important conceptual moves were the introduction of the competence and performance distinction and the E- and I-language distinction. Chomsky insisted that the object of linguistic study should be the physical parts of biological brains that constitute language (I-language) and he suggested that the focus of non-Chomskyan linguists on E-language is misguided. For him E-language is an arbitrary, artificial construct "that is understood independently of the properties of the mind/brain" (Chomsky, 1986a, p. 29), and the study of E-language will not reveal anything interesting about the nature of language.

In the 1990s a further sweeping reconceptualization of the Chomskyan system greatly reduced the number and complexity of required rules and eliminated

deep structure. The resulting Minimalist Program is based on the assumption that syntax is a computational system that provides the optimal solution to the problem of relating sound and meaning. This proposal has been severely criticized, even by theorists closely associated with Chomsky's earlier work (e.g., Culicover, 1999; Jackendoff, 2002; Jackendoff & Culicover, 2005; Newmeyer, 2008), and I will show why these critics are justified.

If we conceive of language as well-defined part of the physical brain, then the brain should be the main object of linguistic study. However, I will show that Chomsky's own research to date has not contributed directly to locating language in the human brain. Further, we shall see that because Chomskyans continue to perceive of empiricist linguistics as behaviourist dogmatism they essentially ignore the results of these researchers. Critics have suggested that this attitude leads to a "time wasting rediscovery of facts or principles that had long been common knowledge outside the Chomskyan camp" (Sampson, 1980, p. 160). My inquiry reveals that to date we have virtually no evidence suggesting that Chomsky's work has contributed directly to locating language in the brain.

It will also become evident that the conceptual move from 'language' defined as set of sentences or expressions to 'language' as part of human brains is problematic. Essentially, this move has never been fully completed and Chomsky continues to treat language as both: as sets of sentences and as biological object (e.g., Chomsky 1986 a, 1995, 2000 a,b, 2007b). This inconsistent treatment blurs the distinction between the object of linguistic study (sentences of a language and their logical relations) and the object of physiological/neurological study (brain

structures involved in generating the sentences linguists can analyze). Chomsky's conflation of the physical tokens of sentences with the non-physical types of sentences results in the untenable view that languages are both finite (as parts of finite brains) and infinite (as grammatical strings of words). It may seem that on a charitable interpretation this dilemma dissolves, if we consider the language faculty as a biological part of the brain that produces a (potentially infinite) set of linguistic expressions. I will show that this interpretation raises different problems for Chomsky's account.

While some important details of Chomsky's theories have changed over the years, he has remained consistent in his core assumptions about language acquisition. As in 1966, in 2009 Chomsky claims that the learning of a language and the resulting linguistic performance cannot be explained adequately by the experience of the language-learning child. He continues to use Poverty of the Stimulus Arguments as a crucial component of support for his postulation of an innate domain specific language acquisition device (LAD). This LAD supposedly is a largely genetically determined part of our biological endowment. Chomsky continues to hold that many aspects of the linguistic performance of an adult speaker cannot be explained by a data-driven general-purpose learning mechanism. I will discuss in some detail the Poverty of the Stimulus Arguments and conclude that we need to pay more attention to empirical data before we can evaluate the soundness of these arguments. Some of these empirical data will be discussed in chapters 4 and 5.

Further, I will provide a detailed account of Chomsky's definitions of the

language faculty and show that after decades of theorizing we still have no precise account of this crucial part of Chomsky's theories. There is no shortage of general descriptions and tentative hypotheses but Chomsky still needs to provide a coherent hypothesis that could be experimentally confirmed or falsified.

My literature review will reveal that the nature of the LAD is still shrouded in mystery. This may seem surprising because for Chomskyans the LAD can "provide an implicit definition of the notion 'human language'" (Allen & van Buren, 1971, p. 14), and a lot of conceptual work has gone into explaining and adjusting it. According to Chomsky his ideas regarding the study of the LAD "crystallized into a distinctive approach to the topic by 1980. In the years since many specific variants have been developed and explored" (Chomsky, 1995, p. 13). However, the frequent re-evaluations of prior variants and wholesale reconceptualizations of previous theories have resulted so frequently in "substantially different conceptions of the mechanisms of language" (Ibid., p. 219) that it has become increasingly difficult to evaluate Chomsky's theoretical commitments at a given point in time. To date Chomsky's work has not provided an unambiguous hypothesis that can be empirically tested.

We will see that some of the main concepts of Chomsky's theorizing are still not well defined. I will highlight this problem discussing the example of 'innateness' and show that (i) Chomsky's own work has not contributed to clarifying this important issue, and (ii) his own use of the term 'innate' is inconsistent and has frequently misled his own followers and his critics alike. I will specifically discuss Chomsky's repeated claims that he has never defended an

innateness hypothesis. I suggest that these claims are misleading and should be replaced by clear statements of the current hypothesis. This will allow to evaluate whether or not this hypothesis is viable.

Next I will show that even though Chomsky's theorizing has evolved over the decades, his commitment to accounting for empirical data has seemingly waned. More precisely, he and some of his followers have become more selective regarding the subset of empirical data they consider to be acceptable for linguistic theorizing. His wholesale dismissal of data gathered by researchers outside his own school is never explained based on studies that have exhibited problematic methods or produced unrepeatable results. In this context I will discuss an unfortunate aspect of Chomskyan science: the many imprecise formulations and contradictory statements. It will become evident that Chomsky's response to criticism is often diverting from the real issues. He does this either by exploiting technicalities that are irrelevant to the issues at hand or by accusing critics of misrepresenting his position without clarifying what his position is. As a result of these tactics many criticisms posed decades ago have still not been adequately addressed. In my opinion Chomsky needs to provide a clear account of his position and of his contributions to linguistics. Until such an account is provided, it is not possible to evaluate whether his contributions have been substantial.

Initially, when Chomsky responded to Skinnerian behaviourism, he critiqued a fairly well-defined set of empirical commitments. Specifically, he rejected the belief that language learning can be accounted for by stimulus-response learning. Even though psychological theories of learning have changed considerably since the

decline of behaviourism (for reviews see Elman et al., 1996; MacWhinney, 2004; Boden, 2006; Edelman & Waterfall, 2007), the arguments of the three editions of *Cartesian Linguistics* do not address these changes. Chomsky's work has been reprinted virtually unchanged, which seems to imply he still holds the beliefs of 1966. In the introduction to the 2009 edition, McGilvray explicitly repeats the earlier claims that "empiricists seem to have added little [to the study of language since Locke] ... like Locke's efforts, theirs generally fail to meet the conditions of adequacy of a *naturalistic* theory" (p. 20, original emphasis). McGilvray claims that this alleged failure is not merely contingent but necessary because the empiricist models of language acquisition are inadequate. He suggests, "with common sense concepts, and especially language, there is no reason to take empiricist speculations at all seriously" (p. 23). Further, he alleges: "No one finds children subjected to the training procedures for concepts or language explored by connectionists, for example" (Ibid.).

In chapters 4 and 5 I will discuss empirical evidence suggesting that, contrary to McGilvray's claims, experimental and computational language acquisition researchers are paying close attention to the conditions under which children acquire language. We will see that many aspects of language acquisition (e.g., word segmentation, acquisition of grammatical categories, past-tense formation, auxiliary fronting in question formation, etc.) are under intense empirical investigation. Researchers work with young children and attempt to develop computational models that simulate the performance of children. This 'empiricist work' will be the focus of the second part of my thesis. In chapter 4 I will introduce

the results of some of the work that has been completed by developmental psychologists, and in chapter 5 I will focus on computational models of language acquisition that are informed by results obtained from the work with children.

In chapter 4 I will focus mainly on the first steps of language acquisition that have been largely neglected by Chomsky's research. Children need to master many cognitive skills in order to acquire and use language. Several of these skills need to be in place long before children begin producing the grammatically complex utterances that are often the focus of Chomsky's work. One of these skills is the ability to produce the sounds of their native language and to combine them into words and eventually into grammatically correct sentences. I will show that it takes a considerable amount of learning before children can reliably produce recognizable words. These learning processes occur over several months and set the stage for later learning. Yet, they are virtually neglected in the Chomskyan approach. In chapter 4 I highlight some of the abilities that the young language learner has to acquire before she can produce her first meaningful sentences.

I will discuss in some detail the stages that precede the production of single and multi-word utterances. In first month of life the infant goes through a phase of vocalization during which she identifies, acquires, and practices the sounds that are common in her language. This babbling stage lasts several months. Around the first birthday, most infants speak their first meaningful words, and they gradually expand their productive vocabulary. Empirical research has shown that the initial pace of vocabulary learning (from birth to 18 months) is very modest. It has been suggested that during this time children acquire and practice many cognitive abilities. An

infant needs to be able to *see* object boundaries before she can form the hypothesis that ostensive definitions apply to whole objects. She needs to be able to *perceive* similarities and differences between objects before she can categorize them. Further, she needs to be able to resolve the conflict between the mutual exclusivity assumption (one name for one object, e.g. 'dog' for the family pet) and the need for taxonomic categorization (e.g. 'dog' for any dog-like object). Children acquire and practice these abilities over an extended time period. Gradually they learn to categorize the world and to understand how words refer to objects, actions, and properties. One hypothesis suggests that once the child has acquired this knowledge, she can slot with ease new words into existing categories (Deacon, 1997). According to this view, general learning mechanisms could account for language acquisition, and an LAD would not be needed.

The fast acquisition of vocabulary and syntax after the second birthday (vocabulary spurt) is frequently used as supporting evidence for the existence of language-specific learning mechanisms that mature at genetically predetermined times (e.g., Chomsky, 1975a, 1985; Lightfoot, 1989; Pinker, 1994; Smith, 1999). I will discuss recent work that offers an alternative account for the vocabulary spurt. On this view the vocabulary spurt is an inevitable result of the infant's immersion in words of varying difficulty, not evidence for the existence of an innate language faculty that is shared by all members of the human species. Further work has shown that not all children go through a well-defined vocabulary spurt. And in cases where a vocabulary spurt occurs its timing varies widely between individual children. These findings suggest that the vocabulary spurt should not be considered as

evidence for genetically predetermined stages of language acquisition.

As children expand their vocabulary, they begin to construct syntactically complex multiword utterances. However, genuine syntactic progress and creativity in language use continue to be rather modest even after the child has acquired a sizable vocabulary. This suggests that the creativity in early language "could be at least partially based upon entrenched schemas and a small number of simple operations to modify them" (Lieven et al., 2003, p. 333). Around 18- to 24- months of age children begin to combine words into coherent utterances. While it is difficult to uncover how children learn the main grammatical categories of verb, noun, adverb, and adjective, there are some indications as to how they do so. I will discuss that children can rely on statistical regularities of the input to uncover crucial information about language from word boundaries over grammatical categories to sentence structure.

Another empirical observation that seemingly supports Chomsky's Poverty of the Stimulus Argument is that certain data (or linguistic constructions such as wh-fronting) appear to be so infrequent in the input that they are virtually inaccessible to the child. This raises the question of whether or not the use of such constructions requires innate knowledge. To answer this question, we need to find out how much (or how little) input is needed to account for data-driven language acquisition. It needs to be determined whether or not all aspects of language competence require the same amount of input. In recent years researchers have begun to analyze the frequency of some crucial constructions in the input (Sampson, 2002). Considering that not all children are exposed to the same data input, it is

important to analyze large samples of data from different backgrounds. This will give us the necessary data to evaluate *appropriately* the competing hypotheses (data-driven learning vs. learning driven by innate knowledge).

Another important issue is explaining the alleged absence of certain kinds of mistakes in children's speech, even though one would expect those mistakes to occur (Chomsky, 1985, 1988). It has been suggested that language comprehension precedes language production and that children do receive feedback when they make comprehension mistakes (overview in Johansson, 2005). I suggest that more data are needed to rule out the possibility that children have learned facts about language, that help them to avoid these kinds of mistakes, at earlier stages of the language acquisition process. In this context it is also important to analyze the mistakes that children *do* make and the kinds of utterances they *fail to produce*, especially in the very early stages of language acquisition. Furthermore, I will discuss the controversial issue of negative evidence (explicit correction of mistakes) in the language input. Chomsky (1977) claims that negative evidence is virtually unavailable. However, some commentators have suggested that different forms of negative evidence are available to the child throughout the language acquisition process. In addition, we will see that negative evidence is not necessary for learning to take place (e.g., Smeeton et al., 2005). Thus, the issue of negative evidence appears to be less clear than implied by Chomsky (1977), and further empirical research is needed to determine whether (and to what degree) negative evidence is required for language learning.

The empirical evidence discussed in chapter 4 does not provide strong

support for Chomskyan nativism/rationalism. Once we analyze carefully all the information that is contained in the language input and account for all steps involved in language acquisition and for the power of general-purpose learning mechanisms, it will become evident that a domain specific LAD may not be *necessary* for language acquisition. But these findings do not rule out that a domain specific LAD exists. As long as Chomskyans do not provide a specific innateness hypothesis that makes testable predictions, empirical research can neither confirm nor disconfirm that a Chomskyan LAD exists.

One way to address the Poverty of the Stimulus Arguments is empirical work with children. Another way is computational modeling of language acquisition. Computational models of language acquisition can challenge the Chomskyan dictum that language acquisition is domain specific and depends on innate knowledge in two different ways. First, if it can be shown that mechanisms that clearly are not involved in human language acquisition can achieve human-like performance, then this proves that the mechanism postulated by Chomsky is not necessary for language acquisition. Second, if connectionist and/or other computational models succeed in simulating language acquisition, then this may allow drawing inferences regarding the nature of the mechanisms that *are* used in human language acquisition.

In chapter 5 I will discuss some of this work and evaluate whether or not McGilvray's claims about the inadequacy of those models are justified. McGilvray claims that the models used by connectionists (the only computational models he considers) are completely inadequate. He provides several reasons for this

inadequacy. First, he holds that "[Connectionists'] claim that the mind is made up of 'neural nets' is innocuous; it is their claim about the initial state of the net (undifferentiated, approximating Locke's 'blank slate') and their view about how this net gets its 'content' (by training, learning) that place them firmly in the empiricist camp" (p. 110). This claim concerns the initial state of the models. Obviously, the 'blank slate' view is indeed problematic. Notably, McGilvray does not provide any examples of connectionists and/or empiricists who hold such an extreme view. My survey of recent literature has found no evidence for such positions. It reveals, instead, that several researchers have explicitly or implicitly rejected completely unconstrained 'blank slate' views of language acquisition (e.g., Hare & Elman, 1995; Elman et al., 1996; Redington & Chater, 1998; MacWhinney, 2000; McDermott, 2001; Solan et al, 2005; Edelman & Waterfall, 2007; Christiansen & Chater, 2008; Chater & Christiansen, 2009).

Second, McGilvray claims: "No one finds children subjected to the training procedures for concepts or language explored by connectionists, for example" (McGilvray, 2009, p. 23). Again, McGilvray does not provide reference to specific work completed by 'connectionists'. I will discuss several examples showing that computational language acquisition researchers are in fact paying close attention to the conditions under which children acquire language. Many computational language acquisition researchers use as input samples of child-directed speech that has been collected by developmental psychologists. The CHILDES database (MacWhinney, 1985) in particular provides a rich resource for computational modeling. Many researchers use samples from this database as input for their

models. Further, several computational models have successfully simulated the performance of language-learning children not only in respect to their successes but also in respect to limitations (e.g., processing of higher order recursion). This may indicate that the mechanisms used in modeling are similar to those used by children. Finally, I will show that many current computational models directly incorporate insights from previous models and from experiments performed with children. Thus, the blanket criticisms of Chomskyans do not apply to all computational work, and McGilvray's claim that computational work makes no contribution to the sciences of the mind should be rejected. The question is not if this work contributes to the sciences of the mind but how it contributes and how these contributions can help to formulate questions for future research. In order to answer this question, individual models need to be evaluated critically.

In chapter 5 I will also show that some of the actual computer-simulations have the potential to provide a model of language learning that does not depend on domain specific mechanisms. However, currently implemented models only simulate small parts of the language-learning task and can, at best, provide evidence that some limited aspects of language learning can be accomplished by non-domain-specific mechanisms. Whether or not the same would hold true for the learning of a complete language 'from scratch' remains to be seen. On the other hand, a failure of attempts to simulate all aspects of language learning on a computer does not necessarily imply that data-driven general-purpose mechanisms are not up to the task. It is possible that the limitations on processing power of currently available computers are prohibitive, particularly in light of the potential need to simulate all

aspects of the child's experience.

The complexity of language has led several authors to conclude that it is implausible that distributional information of patterns within language could play a significant role in the acquisition of syntactic categories (Chomsky, 1975d; Crain 1991; Crain & Pietroski, 2002, McGilvray, 2005). However, it has been demonstrated that a considerable amount of information concerning syntactic categories can be obtained from stochastic information. Natural languages contain rich statistical information and children have powerful learning mechanisms to access this information. The ability to track items based on their perceptual properties allows infants to categorize their natural environment long before they have access to the meaning of words. Recent computational models simulate how children might master the first steps of language acquisition. I will discuss examples of models that succeed in word-segmentation and acquisition of grammatical categories.

I will show that researchers have used computational models to simulate many aspects of the performance of language-learning children. It is important to remember that any successful simulation does not entail that human children acquire language using the same mechanisms. Thus, the computational work cannot disprove the LAD hypothesis. But it can refute the claim that the language input is too impoverished to allow for a data driven general-purpose language learning mechanism. I will provide an overview of recent studies that show how simple recurrent networks (SRN's) and other computational devices can access and use the multiple statistical cues contained in natural languages.

To date computer simulations of many aspects of language acquisition have been performed successfully. I will discuss models that have shown that (i) simple computational models can achieve high accuracy in the word-segmentation task based on positive evidence alone, (ii) correct auxiliary fronting (AUX) in polar interrogatives can be acquired based on indirect statistical information contained in child-directed speech, (iii) "natural" and/or "grammatical" languages are easier to learn than "unnatural" and/or "ungrammatical" languages, (iv) postulating specific mechanisms for rule-based abstraction is not necessary to account for the acquisition of complex grammatical and orthographic rules.

When SRNs and other computational models are able to acquire statistical "knowledge" of the input based on positive examples alone, then it seems to be at least imaginable that children can pick up this information as well. Whether or not children rely on similar mechanisms as SRNs remains a point of debate (for some critical suggestions see Marcus, 1999; Marcus & Brent, 2003). But the success of computational models relying on these mechanisms casts some doubt on the claim that only an innate, domain-specific mechanism can underwrite human language acquisition.

In the final chapter I will summarize the key findings of the previous chapters and provide recommendations for future research. The findings of my research in regards to Descartes' linguistic views cast important doubt on Chomsky's claim that his work can be traced back to a single 'rationalist' Cartesian tradition. This is an important point for the accurate discussion of the history of ideas. But it is only of marginal importance for the contemporary linguistic debates.

Whether or not Chomsky's views can be confirmed by empirical research remains to be seen. However, in spite of the many thorough theoretical reformulations (e.g., Chomsky 1980, 1986b, 1995, 2002, 2005) of these views over the past decades, I can not confirm that Chomsky's "theories have made considerable progress" (McGilvray, 2009, p. 19), if progress is defined in terms of making better empirically testable predictions. Chomsky's work has not clarified the subject of linguistic inquiry, and we still await unambiguous definitions of essential concepts relevant to linguistic research (e.g., 'innate', 'Universal Grammar' 'Language Acquisition Device'). I was also not able to locate specific details about Chomsky's "theory of a biophysically based organic system" (Chomsky, 2010a, p. 20) that could be confirmed or disconfirmed experimentally. The recent proposal that we should not expect that Chomsky's work follows 'normal scientific procedure' (Fiengo, 2006, p. 471), requires clarification in regard to the commitments particular to Chomskyan science. Thus, a main task for future work is to make good on the promises of the 1980s and provide "a principled theory of UG" (Chomsky, 1986b, p. 10) that is descriptively and explanatorily adequate and "accounts for the fact that knowledge of language is acquired on the basis of the evidence available" (Ibid.). Given the vast literature that has been created by Chomskyans, it is essential that they take stock and provide a current overview of the views that have been rejected and the views that are currently held. No such overview has been completed to date.

Regarding the empirical work discussed in chapters 4 and 5, I will suggest that several important tasks remain to be solved. While language acquisition studies can help to cast some doubt on the necessity of domain-specific learning

mechanisms and Universal Grammar, we still know little about the brain-mechanisms that allow children to acquire language. Direct research on human subjects is complicated for a variety of reasons. First, ethical considerations prohibit direct input studies under strictly controlled experimental conditions. Therefore, researchers have to rely on participants who bring an unknown amount of learned knowledge to the experiment, and they "never know for certain what the child has and has not heard [before]" (Tomasello, 2000, p.215). It remains to be seen whether it will be possible to develop methodologies that can distinguish between previously learned and innate knowledge. Second, existing empirical studies tend to focus on very narrow questions. This results in a wealth of empirical data that seem to suggest that certain aspects of language learning can be accomplished by data-driven general-purpose learning mechanisms. But even if we could demonstrate that all of the 'individual pieces' of language competence can be acquired in this way, this does not prove that a domain-specific mechanism is not required to combine and coordinate such complex learning. This is one important question to be examined in future research. Finally, given the tremendous quantity of input available to young children, it would take a substantial amount of time to collect representative samples of even small parts of the actual input. Tomasello cautions that "if we do not know what children have and have not heard, adult like production and comprehension of language is not diagnostic of the underlying processes involved" (Tomasello, 2000, p. 216). Representative input samples will be needed to evaluate adequately the poverty or richness of the input. Keeping the limitations of empirical research in mind, it becomes evident that we are in need of

clear conceptual models that can be tested empirically.

Empirical studies usually address only one small, well-defined aspect of a complex phenomenon like language acquisition. We need to keep this in mind when we evaluate these studies in a broader context. Furthermore, we need to remain open to all possible interpretations of individual results. As Cowie (1999) remarks, empirical data gathered with children cannot rule out the existence of a LAD. It remains possible that a process that appears to rely on a general-purpose learning mechanism could rely on a domain-specific mechanism. As a result, empirical data can only cast doubt on the *necessity* of a LAD. In this context computer-simulations are a particularly helpful tool because they have the potential to provide a model of language learning that does not depend on domain specific mechanisms. However, currently implemented models only simulate small parts of the language-learning task and can, at best, provide evidence that some limited aspects of language learning can be accomplished by domain-general mechanisms. Whether or not the same would hold true for the learning of a complete language 'from scratch' remains to be seen.

Overall, it is essential to strive for conceptual clarity that provides a solid foundation for multidisciplinary research that compares language to more general cognitive capacities in humans and animals. As philosophers we can contribute to this important conceptual work. To do this effectively we need to communicate with researchers who work on the different aspects of language acquisition. Their results need to inform our theorizing. A detailed analysis and evaluation of the problems encountered by Chomskyan linguistics can provide valuable insights for future

theorizing. Such analysis can help to eliminate incorrect assumptions and to question conclusions drawn from them. Such a process is quite Cartesian in spirit: "…[a man] should resolve once and for all to remove from his imagination all traces of the imperfect ideas which have been engraved there up until that time. Then he should begin in earnest to form new ideas, applying all the strength of his intellect" (CSM II, p. 406).

Chapter 2. Cartesian Linguistics

Chomsky (1965a, 1966, 1967, 1975b, 1980, 1985, 1988, 2002, 2007a, 2009b, 2010a) claims that his views of language acquisition are firmly rooted in the Cartesian tradition. This claim has been challenged by Wells (1968), Abelson (1968), Zimmer, (1968), Lakoff (1969), Miel (1969), Salmon, 1969; Aarsleff (1970, 1971), Cohen, 1970; Percival (1972), Barnes (1972), Joly (1985), Seuren (1998), Trabant (2003), Boden (2006), Behme (2009, 2010, 2011), Joseph (2010) and accepted or defended by Katz (1971), Markova (1982), Bracken (1984), D'Agostino (1986), Sorell (1987), Leiber (1988), Barsky (1997), Lightfoot (1999), Smith (1999), McGilvray (2002, 2009), Alanen (2003), Gaukroger (2008). Chomsky insists that his critics misunderstand him and that his interpretation of Descartes and other key figures of the Cartesian tradition is justified (e.g., Chomsky, 1975b, 1980, 1995, 2009a, 2010a). Here I will analyze and evaluate Chomsky's claim regarding Descartes. First I will outline the major points of his view of language acquisition and then inquire whether or not Descartes was committed to similar views.

Chomsky's views have changed considerably over the years, but the following three claims have remained constant and seem to form the essence of Chomskyan Linguistics. First, language is species specific: humans are the only species that has language, and there are no analogues to language in any non-human communication system. Second, language is domain specific. While language is embedded in the broader context of general intelligence, there are certain features of language (e.g., unbounded creativity, recursivity) that are not shared by any other cognitive domain.

This roughly equals what McGilvray (2002) calls 'linguistic creativity'. Third, language cannot be learned from the evidence available to the language learner (Poverty of the Stimulus) and requires therefore some form of innate knowledge or 'innateness' in McGilvray's terminology.

Chomsky published *Cartesian Linguistics* with the purpose of deepening "our understanding of the nature of language and the mental processes and structures that underlies its use and acquisition" (Chomsky, 1966, p. ix). He suggested that this could be partly achieved by re-analyzing the contributions of an earlier European tradition. Chomsky claims that during the Cartesian period there was "a coherent and fruitful development of a body of ideas and conclusions regarding the nature of language (Ibid., pp. 2-3) and promises to "determine the exact nature of this 'capital of ideas' … [and] to evaluate the contemporary significance of this contribution" (Ibid., p. 3). These are undoubtedly laudable goals. An appreciation of these earlier contributions would presumably not only benefit linguistics but also make important contributions to our understanding of historical and philosophical issues. The importance of historic texts for contemporary research in philosophy has been clearly acknowledged: "Doing systematic philosophy is part of doing history of philosophy, and doing history of philosophy is part of doing systematic philosophy" (Katz, 1986, p. 5). Thus, Chomsky's suggestion to adopt a similar approach to linguistics might encourage multi-disciplinary research.

In the 1966 edition of *Cartesian Linguistics* we find an attempt to draw out important insights from philosophers and linguists from the seventeenth, eighteenth and early nineteenth centuries, "the period of 'Cartesian linguistics'" (Chomsky,

1966, p. 1) and relate them to the theorizing Chomsky had pursued since the 1950s. Chomsky acknowledges that a full account of the ideas of the Cartesian period would go beyond the scope of his essay, which he calls "a preliminary and fragmentary sketch of some of the leading ideas of Cartesian linguistics" (Chomsky, 1966, p. 2). Nevertheless he seems confident that he can show that during this period there was "a coherent and fruitful development of a body of ideas and conclusions regarding the nature of language in association with a certain theory of mind" (Chomsky, 1966, pp. 2-3).

Of special interest to Chomsky's project are 4 points that he attempts to connect to Cartesian theorizing. He dedicates a chapter to each of these points: (i) the creative aspect of language use (pp. 3-30), (ii) the distinction between deep and surface structure of language (pp. 31-51), (iii) the use of description and explanation in linguistics (pp. 52-58), and (iv) the acquisition and use of language (pp. 59-71). As my focus in this chapter will be on the contribution of Descartes my evaluation will focus mainly on (i), (ii), and (iv). This seems in spirit with Chomsky's project. He clearly recognizes that "Descartes himself devoted little attention to language" (Chomsky, 1966, p. 2). But he insists that the term 'Cartesian' is nevertheless justified because of the contributions Descartes made to a theory of mind Chomsky calls "an outgrowth of the Cartesian revolution (Chomsky, 1966, p. 3).

The creative aspect of language use was a central concern of Chomsky's early work in linguistics. He opposed the then dominating behavourist views of language (e.g., Chomsky 1959, 1964) and attempted to demonstrate that the fundamental difference between human language and all known non-human communication

systems was the ability of humans to use language stimulus free, appropriately and innovatively. Citing several passages from Descartes' writing he showed that there were important similarities between his own views and Descartes' view:

> In summary, it is the diversity of human behaviour, its appropriateness to new situations, and man's capacity to innovate-the creative aspect of language use providing the principle indication of this-that leads Descartes to attribute possession of mind to other humans, since he regards this capacity as beyond the limitations of any imaginable mechanism. Thus a fully adequate psychology requires the postulation of a 'creative principle' alongside the 'mechanical' principle that suffices to account for all aspects of the inanimate and animate world and for a significant range of human actions and 'passions' as well. (Chomsky, 1966, p. 6)

I will discuss additional texts that support the claim of Chomsky that Descartes shared his view that language required an explanation that goes beyond mechanistic principles. But it will also become evident that there are important differences between Chomsky's and Descartes' view regarding the nature of the 'creative principle'. When looking for historic antecedents to current theorizing, Chomsky does not only discuss Descartes' and other Cartesians' writings that support his own view. He also contrasts the Cartesian view to that of thinkers (e.g., LaMettrie, Bougeant) who held language does not indicate a fundamental difference between humans and animals because the actions of both could be fully explained based on mechanistic principles alone. According to Chomsky these authors are mistaken because they fail to

> …come to grasp with the problem raised by Descartes-the problem posed by the creative aspect of language use, by the fact that human language, being free from control by identifiable external stimuli or internal psychological states, can serve as a general instrument of thought and self

expression rather than merely a communicative device of report, request or command. (Chomsky, 1966, pp. 11-12)

Chomsky attempts to tie these perceived historic antecedents to the view of contemporary theorizers: "Modern linguistics has also failed to deal with the Cartesian observations regarding human language in any serious way (Chomsky, 1966, p. 12). This is an allegation that may have been justified in 1966 (e.g., for Ryle, Bloomfield), but, I content, it does not reflect the situation in 2009. Yet, the formulation 'modern linguistics' has been re-printed unchanged in the 2nd and 3rd edition of *Cartesian Linguistics*. I will discuss McGilvray's failure to account for changes in the intellectual climate against which Chomsky's works should be read in detail later. For now I shall focus on Chomsky's original contribution.

Chomsky attempts to tie the Cartesian insights regarding the creative use of language to "the substantive discussion of grammar …[first provided by the] …Port-Royal *Gramma*r" (Chomsky, 1966, p. 31). He argues that proponents of the Port-Royal Grammar were committed to a view that is very similar to his own. In particular he sees them arguing for the same kind of deep structure/surface structure divide he is proposing. This divide is based on Cartesian dualism: "…a sentence has an inner *mental aspect* (a deep structure that conveys its meaning) and an outer, *physical aspect* as a sound sequence" (Chomsky, 1966, p. 40, emphasis added). And, as in Chomsky's own theorizing transformational rules convert deep to surface structure and recursive devices allow, "for infinite use of finite means…as any adequate theory of language must" (Chomsky, 1966, p. 41). Chomsky further claims that the Port-Royal *Logic* was inspired by Descartes' system of ideas: "… the

[Cartesian] usage of the term idea as anything that can be that can be 'conceived' (not merely imagined'), is the one carried over to the Port-Royal *Logic*. In this sense, concepts of varied types, even propositions are ideas" (Chomsky, 1966, p. 98). An adequate technical discussion of all the similarities that Chomsky perceives between the Port-Royal *Grammar* and *Logic* would lead us too far afield. Therefore I will restrict the discussion to a detailed analysis of Cartesian ideas in sections 2. 4. I will argue that in spite of some similarities between Descartes' and Chomsky's conception of ideas there are also very important differences that need to be more adequately acknowledged than Chomsky has done.

In the section on language acquisition Chomsky proposes that the "central doctrine of Cartesian linguistics is that the general features of grammatical structure are common to all languages and reflect certain fundamental properties of the mind" (Chomsky, 1966, p. 59).

> …there are certain language universals …[that] are not learned; rather they provide the organizing principles that make language learning possible, that must exist if data is to lead to knowledge. By attributing such principles to the mind, as an innate property it becomes possible to account for the quite obvious fact that the speaker of a language knows a great deal that he has not learned (Chomsky, 1966, pp. 59-60)

Chomsky suggests that this focus on linguistic universals also can be traced back to the work of Descartes. He claims that, "the Cartesian origins of a 'gammaire générale'…and a 'grammaire raisonnée'…are too obvious to require discussion" (Chomsky, 1966, p. 106) and suggests a connection to Descartes' theory of cognition as discussed in the Comments on a Certain Broadsheet (CSM I, 304). Chomsky claims that later rationalist thinkers were inspired by this work when they attributed

innate resources to language acquisition. He discusses specifically Herbert of Cherbury's theory of common (innate) notions (Ibid., p. 60), Cordemoy's theory of language acquisition (Ibid., p. 63), Humboldt's Platonistic claim 'die Erlernung ist…immer nur Wiedererzeugung' and postulation of a critical period of intellectual development (Ibid., p. 64) and Cudworth's attempts to unify a theory of perception with a theory of learning (Ibid., p. 66). Chomsky draws repeatedly direct connections between the work of later thinkers and Descartes. For example Chomsky suggests that, "Herbert expressed much of the psychological theory that underlies Cartesian linguistics, just as he emphasized those aspects of cognition that were developed by Descartes" (Ibid, p. 62) and after citing Descartes 'theory of cognition' from the *Comments on a Certain Broadsheet* he claims "Rather similar ideas are developed at length by Cudworth" (Ibid., p. 67). We see here clearly an effort by Chomsky to link the work of a later period directly to Descartes' writings. This later work in turn is seen as important predecessor of contemporary research in language acquisition, sense perception and psychology in general with a focus on the role of innate schemata and a system of fixed rules generating more complex schemata. Chomsky concludes, "…it would be quite accurate to describe current work as a continuation of the tradition of Cartesian linguistics and the psychology that underlies it" (Chomsky, 1966, p. 72).

In the years following the first edition of *Cartesian Linguistics* it has been questioned just how accurate Chomsky's description of the historic antecedents of his own work was. Several scholars (e.g., Zimmer, 1968; Lakoff, 1969; Salmon, 1969; Aarsleff, 1970, 1971; Percival, 1972) have expressed dissatisfaction with the historic

accuracy of the account given by Chomsky and with the at times polemical style of the exposition. For example, the claim that Cartesian linguistics is Cartesian because its proponents "all hold, in some degree, views … emanating from Descartes, however remotely" (Salmon, 1969, p. 165) has been shown to be false (e.g., Salmon, 1969; Lakoff, 1969; Percival, 1972). To be clear, the allegation is not that Chomsky overlooked some insignificant historical detail in an otherwise accurate depiction of the rationalist tradition. According to historians the core claims that tie Chomsky's own theorizing to earlier rationalist scholars have been questioned and it has been suggested that the term 'Cartesian Linguistics' is "thoroughly misleading" (Percival, 1972, p. 144).

Salmon has focused on "three points, which Chomsky regards as essential in postulating a relationship between Descartes and the tradition of universal grammar which, as he says, 'develop from the Port-Royal *Grammaire générale*'" (Salmon, 1969, p. 169). These points concern (i) the concept of universal grammatical principles, (ii) the concept of deep and surface structure and (iii), the existence of innate ideas that allow for language acquisition. Her detailed analysis of history shows that Chomsky's claims regarding all three points are questionable.

The concept of universal grammatical principles had been already recognized by scholars before Descartes time: "there is ample evidence that there was a full awareness of the opposition between general and special grammar long before any of Descartes' works had been printed" (Salmon, 1969, p. 174). Hence, it is not clear that the Port-Royalists took their inspirations for deepening an understanding of such principles from Descartes or from earlier sources. Similarly, the "realization that deep

and surface structure need not be identical is not necessarily … a Cartesian feature" (Salmson, 1969, p. 175) but had already been recognized by earlier scholars (e.g., Bibliander, 1548; Ramus, 1578; Sanctius, 1587; cited by Salmon, p. 175) who worked in the Aristotelian tradition. The distinction between deep and surface structure was not an original contribution by the Port-Royalists but was "derived from a logical and grammatical tradition which had been developing without any real interruptions since the early Middle Ages" (Salmon, 1969, p. 178). Finally, Chomsky's interpretation of history regarding the acquisition and use of language is problematic. The focus of Chomsky's account is on Descartes' important contribution to a rationalist tradition that holds language acquisition depends crucially on innate resources. This focus distracts from the fact that a debate, not unlike the one between Chomsky and contemporary empiricist/behaviourist linguists, occurred in the early 17th century. The roots of this debate can be traced to the early 16th century when methods for effective teaching of the academic language Latin were hotly debated. The point of contention was whether Latin should be taught by focusing on repetition of patterns or by teaching of rules for forming grammatical sentences. Thus, the subject that became focus of the debates in the 1950s "was one of major importance before any Cartesian could have been felt" (Salmon, 1969, p. 185). Salmon concludes that the main weakness of Cartesian Linguistics "is the attribution of the form taken by the Port-Royal grammar to Cartesian inspiration with little or no attempt to take into account the total intellectual context in which it appeared" (Salmon, 1969, p. 185).

Similar concerns about the accuracy of Chomsky's attempt to relate some

central features of his own linguistic work to alleged Cartesian antecedents have been expressed by other scholars (e.g., Zimmer, 1968; Lakoff, 1969; Aarsleff, 1970, 1971; Percival, 1972; Barnes, 1972). Keith Percival suggests that an intellectual movement as suggested by Chomsky probably did not exist. He questions that Descartes' work made important novel contributions to linguistic theorizing and shows that Chomsky has provided no conclusive evidence supporting the claim that Descartes' work has influenced the Port-Royal grammarians. Further, he shows that there were some important differences between the commitments of Descartes and Chomsky. For Chomsky both animals and humans have intelligence and human language expresses the highest degree of intelligence. For Descartes on the other hand only humans are intelligent and thus the tests he developed for creativity are in essence tests for the presence of *any* intelligence. In *Discourse on Method* Descartes contrasts the limited faculties of animals to human reason and he observes that the latter is "a universal instrument which can be used in all kinds of situations" (CSM I, p. 140). However, language use requires only a bare minimum of intelligence: "it patently requires very little reason to be able to speak" (Ibid.). Descartes goes even further and shows that even madmen (but not parrots) can use language appropriately. This cannot be an indication of intellectual sophistication but merely shows that humans are capable to give answers that are relevant to a subject matter. Percival suggests that, "the notions of innovativeness and potential unboundedness which are basic to Chomsky's view of language play a negligible role in Descartes' discussion of the subject" (Percival, 1972, p. 142). This is so because for Descartes mind, not language is of primary concern in these passages. This point has also been stressed independently: "What

Chomsky did not clearly saw when he interpreted Descartes was that for Descartes it is not language that is creative but the human mind" (Trabant, 2003, p. 137). I will return to this difference in section 2.1. And, like Salmon, Percival comments on the work on syntax that preceded Descartes' and was much more detailed than anything written by Descartes on this subject. Hence it would be reasonable to assume the Port-Royal grammarians relied on this earlier work when they explored the distinctions between general and particular grammars.

Robin Lakoff grants that there might have been genuine antecedents to Chomsky's transformational grammar in the time in question. However, she argues that these antecedents predate the Port-Royal Grammar and cites as one example Lancelot's *Nouvelle methode pour facilement et en peu temps comprendre la langue latine* (NML) which was first published in 1644 (Lakoff, 1969, p. 347). This was a pedagogical grammar based on the idea that all languages share numerous traits "due to the logical mind common to all men-and that by making use of these logical similarities, one could teach Latin better than by forcing students to memorize sentences" (Lakoff, 1969, p. 347). According to Lakoff NML shared indeed some important commitments with Chomsky's linguistic theory but there were also important differences, especially in how deep structure was conceived: "…[NML] goes no further into the logical structure of the mind than possible surface structures of the language allow. This is quite different from most present-day views of deep structure" (Lakoff, 1969, p. 252). Lakoff also notes that we find no acknowledgement of any intellectual debt to either Descartes or Cartesian philosophy in the main works of Port Royal grammarians. Instead, we find explicit acknowledgement of the

important contribution of Sanctius. And Lakoff shows that the work of Sanctius contained indeed important insights that have been incorporated by the Port-Royalists and do bear some resemblance to Chomsky's theories:

> (1) Language is a product of the human mind. Since the mind is a rational thing, so is language.
> (2) It is necessary to look for explanations for grammatical phenomena. These may not always be immediately evident, but that doesn't mean they don't exist.
> (3) Nothing in language is accidental.
> (4) Prescriptivism is worthless: a grammarian's authority is only as valid as the logic behind his analyses. (Lakoff, 1969, p. 359)

It would seem then that these insights were neither 'linguistic innovations' of the Port-Royal school nor were they due to the influence of Cartesian philosophy. Lakoff concludes it would be more appropriate to call the resulting view Sanctian linguistics.

Finally, Hans Aarsleff published two critical reviews in which he echoes some of the concerns already discussed and stressed that these concerns affect the very core of Chomsky's arguments. But he also stresses another important issue that questions whether there was a particularly rationalist linguistic tradition that influenced scholars during the Cartesian and Romantic periods. Aarsleff stresses that not Descartes but John Locke had an important influence on subsequent linguistic theorizing, especially on Condillac, "the most important figure [of 18th century linguistics]" (Aarsleff, 1971, p. 576) and subsequent scholars. According to Aarsleff these scholars have been profoundly influenced by Locke, yet they hold views that Chomsky classifies as Cartesian: " universal grammar of the highest order can be done… on a Lockean basis" (Aarsleff, 1971, p. 580). Given that Locke is usually classified as empiricist a problem arises for Chomsky's rationalist interpretation.

Either there was no distinct Cartesian but rather a Lockean influence on subsequent linguistic theorizing. Or, if we want to salvage the label 'Cartesian', we have to assume that the views on language of Locke and Descartes were not incompatible. But then it would seem odd to insist on the rationalist/empiricist distinction as Chomsky does. Overall, for Aarsleff "Chomsky's version of history is the product of serious deficiencies in knowledge and research, and is an obstacle to the creation of a true and significant history of linguistics" (Aarsleff, 1970, p. 570).

In Chomsky's defense it might be argued that his goal was not to give an entirely accurate of the history of linguistics. Even if *Cartesian Linguistics* is neither a good history book nor an accurate depiction of a singular Cartesian or even rationalist linguistic research strategy it still could have been a valuable contribution. This much has been suggested: "[Chomsky] was just providing a source book for transformational linguists to see how their work manifested important philosophical currents ... Judged as spade work in a neglected (indeed, rejected) area of language philosophy in order to inform current practice, it's very fine indeed, even thrilling" (Harris, 1998). This may have been the case for the 1966 edition. But, it would seem that in order to continue to 'inform current practice' the factual mistakes contained in the original work need to be eliminated. Yet, the 2002 and 2009 editions are virtually unchanged reprints of the 1966 edition. This is regrettable because is has been suggested that one of the main values of Cartesian Linguistics was its attempt to establish historical connections between philosophical and linguistic theorizing:

> ...it is greatly to Chomsky's credit that he has boldly advanced historical hypotheses which more pedestrian scholars would not have had the courage to publish. For in committing what might seem like an academic

indiscretion, Chomsky has revealed the true extend of our present ignorance in this whole area" (Percival, 1972, p. 145).

Finally, Chomsky's project has also been enthusiastically defended against some of his critics. Harry Bracken argued that especially Aarsleff and Lakoff have thoroughly misunderstood Chomsky's intentions, the commitments of Chomskyan linguistics and the relevance and importance of the historical texts. Unfortunately, Bracken's exposition is riddled by personal attacks that culminate in the accusation that "Aarsleff's confusions and misunderstandings are so profound that one is reminded of Malebranche's advice …[When one criticizes a book it seems to me, one should at least have read it]" (Bracken, 1982, p. 115). It is of course possible to disagree with Aarsleff's interpretation of Chomsky's book. But, in light of the countless passages provided by Aarsleff, to suggest he may not even have read *Cartesian Linguistics* raises doubts about Bracken's ability to give an objective evaluation. His dismissal of Lakoff's review is similarly problematic as he tries to establish based on a single quotation that her exposition of historic texts is "thoroughly misleading" (Bracken, 1982, p. 122). The following evaluation may indicate a serious misunderstanding on Bracken's part: "Salmon's 1969 review… is an exemplary model of the kind of good scholarship Chomsky's work has provoked" (Bracken, 1982, p. 121). Given that Salmon agreed on factual matters with the other critics and concluded that *Cartesian Linguistics* was seriously flawed it is surprising that Bracken considers her scholarship 'exemplary' but accuses Aarsleff and Lakoff of incompetence and misrepresentation. Nevertheless, there are two valid points in Bracken's review. He insists that before we argue about the appropriateness of the

43

label 'Cartesian' it "is of primary importance … whether the historian of Cartesian thought gets Descartes right" (Bracken, 1982, p. 114). For this reason I will focus in the main section of this chapter on Descartes' thought and show that Chomsky did not always get it right. In order to do this, I will follow another piece of Bracken's advice: "… the texts of Chomsky, Locke, Descartes, and the Cartesians are available. Serious students should study them and make their own judgments" (Bracken, 1982, p. 123).

As we have seen, even Chomsky's critics agree that it would be a worthwhile exercise to "determine the exact nature of the capital of ideas accumulated in the premodern period" (Chomsky, 1966, p. 2). Chomsky's initial attempt to do this was flawed but, given that he had been made aware of the problems, it would seem natural to expect, that subsequent editions of *Cartesian Linguistics* had eliminated the errors. Yet, the more recent editions only differ from the original by providing English translations of quotes that had been in French and German in the original and by adding lengthy introductions penned by Editor McGilvray. The main text is virtually unchanged from the 1966 original. I believe that Chomsky's defense of his methodology as a historian remains questionable. First, as editor McGilvray explains in a footnote (p. 109f), Chomsky advocates what could be called the 'selective-history-approach' (SHA):

> One might say that I'm looking at history ... from the point of view of …
> an art lover who wants to look at the 17th century to find in it things that
> are of particular value and that obtain part of their value … because of the
> perspective with which he approaches them (Chomsky, 1971).

This passage indicates that for Chomsky it continues to be more important

what he wants to see in the Cartesians than what those people actually thought. And, in case the critics might still complain that art-lovers frequently agree on which pieces by an artist are worth collecting, Chomsky adds later what I would call the 'rewrite-history-approach' (RHA):

> The first [question], the actual sequence of events, is not in itself very interesting in my opinion; it's a story of chance events and personal accidents, accidents of personal history. The second question, namely, *how it should have happened*, is far more interesting and important, and that certainly has never been told or even investigated (Chomsky, 1997, emphasis added).

This passage shows clearly that Chomsky has little interest in the facts of history, but intends to use the suitably re-interpreted Cartesians as figurants or ventriloquist puppets on the Chomskyan-Linguistics stage. Combined, SHA and RHA allow Chomsky to pick and choose what he considers of value in the Cartesians' writings and to transform other passages into 'what the Cartesians should have written'. This might seem to justify the artistic freedom Chomsky applies to history. Using the example of Descartes I will provide several examples showing that the resulting 'hybrid-view' has little resemblance to any view Descartes held. In the process of Chomskyan rewriting of history it has become nearly irrelevant what Descartes said, and while the resulting picture might be interesting in his own rights, it is misleading to call it Cartesian.

Throughout this chapter I will provide textual evidence supporting my claim that, when we allow Descartes to speak on his own behalf, it becomes dubious that Descartes contributed to "a coherent and fruitful development of a body of ideas"

(Chomsky, 1966, pp 2-3) that culminated more than 250 years later in Chomsky's linguistics. While there are some similarities between the views of Descartes and Chomsky, I will show that their differences are considerably greater and that the term *Cartesian Linguistics* is misleading. Specifically, I will show that while Descartes believed that language is one of the distinguishing features between humans and non-human animals, he was not committed to the view that language is domain-specific, nor did he seem to hold that language depended on the kind of innate knowledge Chomsky considers crucial. Because the last claim is the most controversial, I will spend the main part of this chapter providing evidence in support of it. Before we get there I shall deal with the less controversial claims: Descartes held language to be species specific but not domain specific.

2.1. Species Specificity of Language

Many Descartes scholars (e.g., Kenny, 1968; Percival, 1972; Lindeboom, 1978; Rodis-Lewis, 1978; Wilson, 1978, 1995; Bracken, 1982; Joly, 1985; MacDonald Ross, 1988; Cottingham, 1986, 1992; Leiber, 1988; Musgrave, 1993; Seris, 1993; Gaukroger, 1995; Baker & Morris, 1996; Rozemond, 1998, 2006; Morris, 2000; Sutton, 2000; Wright, 2000; Des Chene, 2001; Alanen, 2003; Ariew et al., 2003; Wilson 2003; Clarke, 2003, 2006; Sorell, 2005; Skirry, 2005; Williams, 2005; Brown, 2006; Hoffman, 2007; Schmaltz, 2006; Hatfield, 2008a,b) observe that for Descartes one of the distinguishing characteristics of humans is the use of language: virtually all humans but no animals are able to use language. This observation is well supported by textual evidence. We find numerous comments

related to language and its importance for human thought and communication throughout Descartes' writing (e.g., CSM I, p. 140f, 220, 348; CSM II, p. 128, 198; CSMK III, p. 10-13, 102, 303, 366, 374). Several of these comments address the fundamental differences between humans and non-human animals. For example in *Discourse on Method* Descartes explains: "...there are no men... [who] are incapable of arranging various words together and forming an utterance from them in order to make their thoughts understood; whereas there is no other animal... that can do the like" (CSM I, p. 140). He makes clear that the linguistic differences between man and animal are not based on differences in their bodies (Gunderson, 1964; Percival, 1972; Lindeboom, 1978; Bracken, 1984; Leiber, 1988; Cottingham, 1992, 1997, 2008; Baker & Morris, 1996; Wilson, 2000; Morris, 2000; Gaukroger, 2002; Grush, 2003; Williams, 2005; Gombay, 2007; Hatfield, 2008b).

Descartes holds that on the one hand it would be impossible to distinguish a machine which had the organs and the appearance of a monkey from a real monkey. On the other hand a machine resembling the appearance of a human could be easily distinguished from a real person because the machine "could never use words or other signs for the purpose of communicating its thoughts… [and it were] …not acting through understanding but only from the dispositions of [its] organs" (CSM I, p. 140). Descartes holds that the ability to engage in meaningful conversation derives from our ability to reason, which "is a general instrument which can be used in all kinds of situations" (Ibid.). By contrast, an automaton built by us would be stuck with a series of pre-programmed simple speech dispositions.

Several of Descartes' letters also express his conviction that, regardless of the

many similarities between humans and animals, the use of language sets the former apart from the latter (e.g., 'language test' Gunderson, 1964; Lindeboom, 1978; Joly, 1985; MacDonald Ross, 1988; Leiber, 1988; Cottingham, 1992, 1997, 2008; Baker & Morris, 1996; Wilson, 2000; Morris, 2000; Erion, 2001; Grush, 2003; Williams, 2005; Brown, 2006; Schmaltz, 2006; Hatfield, 2008b). In his *Letter to Reneri* for Pollot Descartes explains that automatons replicating animals and those replicating human could be distinguished because the former "never answer in word or sign, except by chance to questions put to them" (CSMK III, p. 99). In the same letter he cautions against our tendency to infer from a resemblance of behaviour a resemblance of mechanisms underlying the behaviour: "... the resemblance between some exterior actions of animals and our own... is not at all a sufficient basis to prove that there is any correspondence between the corresponding interior actions" (CSMK III, p. 100).

A similar point is made in a *Letter to Newcastl*e. Here Descartes admits that a comparison between most observable forms of behaviour would not reveal fundamental differences among machines, animals and humans. But he insists that "spoken words, or other signs that have reference to particular topics without expressing any passions" (CSMK III, p.303) are what set humans apart from animals and machines; "the use of words, so defined is something peculiar to human beings" (Ibid.). Similarly, in a *Letter to More* he writes, "...it has never been observed that any brute animal has attained the perfection of using real speech, that is to say, of indicating by word or sign something relating to thought alone and not to natural impulse" (CSMK III, p. 366). Descartes carefully highlights the fundamental difference between superficial but misleading similarities in communication

behaviour on the one hand and true language use on the other (Gunderson, 1964; Lindeboom, 1978; MacDonald Ross, 1988; Leiber, 1988; Cottingham, 1992, 1997, 2008; Baker & Morris, 1996; Erion, 2001; Alanen, 2003; Grush, 2003; Williams, 2005; Schmaltz, 2006; Hatfield, 2008b). This is especially important because it accounts for the fact that some animals (e.g., parrots, magpies) have been taught some words and seemingly use these words to the same ends as humans do. Yet, according to Descartes there is a fundamental difference between word use in birds and humans: "If you teach a magpie to say good-day to its mistress when it sees her approach, this can only be by making the utterance of this word the expression of one of its passions. For instance it will be the expression of the hope of eating if it has always been given a tidbit when it says it" (CSMK III, p. 303). Descartes is concerned not with the overt behavior of the animal but with the reasons for this behaviour. The magpie does not say 'good-day' to its mistress because it knows the meaning of the words but because it has learned that it will be rewarded with a tidbit. Therefore the use of the words is not any different from the tricks a dog might perform once it has learned that it will be rewarded for such a performance: "Similarly, all the things which dogs, horses and monkeys are taught to perform are only expressions of their fear, their hope or their joy; and consequently performed without thought (Ibid.). It is important for Descartes to use this direct comparison. This allows him to show that what the magpie has learned is not a very small part of human language but a novel way to express one of its passions.

Maybe the strongest expression of Descartes' conviction that language is species specific is his short reply to More where he states that "Infants are in a

different case from animals: I should not judge that infants were endowed with minds unless I saw that they were of the same nature as adults; but animals never develop to a point where any certain sign of thought can be detected in them" (CSMK III, p. 374). This passage is significant because, unlike in the earlier cases, Descartes cannot provide any empirical evidence for his claim that infants are endowed with minds. Observation would tell us that infants are indistinguishable from animals because they are not able to use language. Nevertheless, Descartes is convinced that their inability to speak is different from that of animals. Animals will never 'develop to a point' where they acquire language, while infants will grow up to become language users. They are 'of the same nature as adults'; they have minds (Cottingham, 1978; Marles, 1978; Hoffman, 1986; Leiber, 1988; Wilson, 1995; Baker & Morris, 1996; Hatfield, 2007). Therefore it is membership in a class (humans), not currently displayed behaviour, that determines whether or not an organism has or will develop the ability to use language. For Descartes it is clear that only humans belong in the class of language users.

In this context I need to discuss an apparent misunderstanding in Chomsky's view of Descartes' commitments. Seemingly Chomsky believes that Descartes denies that animals can use language because they are not as intelligent as humans. He writes:

> Does the inability of other species to develop language of the human type derive from their lack of a specific quality of intelligence rather than from a mere limitation in a common intelligence, as Descartes thought? (Chomsky, 1975a, p. 40)

Even though this sentence is ambiguous, it indicates a problem with Chomsky's interpretation of Descartes. The phrase 'as Descartes thought' refers either to 'lack of a specific quality of intelligence' or to 'a mere limitation in a common intelligence' or to both. However, the text clearly confirms that Descartes believed that animals have no common intelligence at all (see discussion next section) and therefore they are not able to 'develop language of the human type'. So the statement expressed in Chomsky's sentence is incorrect under either possible interpretation. Another passage shows that Chomsky also seems to be unclear about the nature of the fundamental difference that Descartes sees between humans and non-human animals: "the major problem that Descartes and his followers raised about the apparent limitations of machines, …had nothing to do with the acquisition of language and other 'mental systems' but rather with their use, *an entirely different matter*" (Chomsky, 2010b, p. 104, emphasis added). We have already seen that Descartes would deny that animals (or automatons) would be able to acquire language. In the next section I will show that Descartes also did not draw the kind of conceptual distinction between language acquisition and language use Chomsky attributes to him. For Descartes language learning is made possible by the human mind. Only organisms that have minds can learn the words and grammar of a language and use this language creatively. From Chomsky's contemporary perspective it may be sensible to assume that the view that animals lack intelligence entirely was wrong or that there is an important difference between language acquisition and language use. But, in my opinion, this does not justify attributing to Descartes views he did not hold.

In summary, the textual evidence strongly suggests that Descartes believed that language is species specific: all humans have it regardless of a wide range of differences in their health, age, and intelligence. On the other hand, no animal has it. However, the fact that language is species-specific could be explained in two different ways. In the first scenario humans have a domain-specific language faculty while animals lack this faculty. In this case it would be imaginable that an animal, that had implanted an artificial language faculty would behave in ways that are indistinguishable from a human being. This is essentially what Alan Turing (1950) believed when he conceived of the Turing test: if a machine displays language behaviour that is indistinguishable from that of a human speaker, then we would believe that this machine is intelligent. Chomsky seems to be committed to a similar view. In the second scenario language is one of several indicators of general intelligence (thought and reason in Descartes' terminology). In this case it would not be possible to 'insert' a language faculty that is independent of 'general intelligence'. The differences between humans and non-human animals are not merely differences in their ability or inability to use language but more fundamental differences. To bring it to a point: in the first scenario humans are intelligent because they have language while in the second scenario humans have language because they are intelligent. Some commentators hold that Descartes was committed to the first scenario. For example, John Cottingham writes, "Descartes unequivocally advanced the claim that there is *no thought without language*" (Cottingham, 1997, p. 30, original emphasis). In the next section I will provide evidence for my claim that Descartes was committed to the second scenario.

2. 2. Domain Specificity of Language

Chomsky's commitment to a domain specific language faculty is well documented. He stresses that language is learned earlier and with more ease than any other intellectual skill (e.g., mathematics), that it appears to be largely independent of the actual input and the overall intelligence and that it seems to be affected by specific types of brain damage that have no effect on 'overall' intelligence (Chomsky, 1975a, c, 1988, 1995, 2002). Many of Chomsky's followers (e.g., Brook & Stainton, 2000; Cattell, 2006; McGilvray, 1999, 2002; Lightfoot, 1999; 2002; Russell, 2004; Smith, 1999; Stainton, 1996) accept these indicators for a domain specific language faculty.

Several commentators (e.g., Chomsky, 1968, 1975b; Bracken, 1984; Cowie, 1999; McGilvray, 2002) have suggested that the idea of a domain-specific language faculty can be traced back to Descartes' writings. They claim Descartes believes that we have some specialized faculties of the mind that allow the acquisition of ideas that could not have been acquired from experience (e.g., the idea of God, mathematical concepts). I will now inquire whether Descartes was indeed committed to such specialized faculties or whether he held that our faculty of rational thought allows for the acquisition of our ideas (as suggested by Gunderson, 1964; Kenny, 1967; Schmaltz, 1997, 2008; Erion, 2001; Grush, 2003; Williams, 2005). The acquisition and use of language will remain the focus of this inquiry.

At first glance Descartes seems to endorse a domain-specific view of

language; he seems to allow that language use can be independent of intelligence (reason). In a *Letter to Reneri* he remarks that automatons replicating animals attempting to imitate our use of language would "fail more disastrously than the greatest fool" (CSMK III, p. 99). In the *Letter to the Marquess of Newcastle* he refers to "the speech of madmen which has reference to particular topics even though *it does not follow reason"* (CSMK III, p. 303, emphasis added) and in one *Letter to More* he remarks, "...all human beings use [language], however stupid and insane they may be..." (CSMK III, p. 366). In *Discourse on Method* the reference is equally clear: "...there are no men so dull-witted or stupid - and this even includes madmen - that are incapable of arranging various words together and forming an utterance from them..." (CSM I, p. 140). Here as in the letters, the verdict appears to be that language can be sustained without a lot of intelligence: "For it patently requires very little reason to be able to speak" (Ibid.). Seemingly, it requires so little intelligence that "even the stupidest child or ... a child with a defective brain" (Ibid.) can still acquire and use language. However, Descartes claims at no point, that *no* intelligence (reason) is required for language. Nor does he suggest anywhere that language depends on a specific quality of intelligence. This is important because Descartes denies that any human being could be entirely incapable of using language (CSM I, p. 140, CSMK III, pp. 303-304). However, Descartes also suggests nowhere that there is no qualitative difference between the language of a dull-witted fool and that of a 'normal' person. So the question arises how language and intelligence are connected on Descartes' view.

When we pay close attention to context, we recognize that the purpose of

Descartes' examples is to highlight what he takes to be the fundamental difference between humans and animals; the possession or lack of a mind (intelligence). So language is not *the* distinguishing feature but rather a reliable indicator for this feature. This view is confirmed by the text. In *Discourse on Method* Descartes discusses the abilities of animals and observes that (some) animals have the necessary organs to speak and show more skill in some tasks than humans do but nevertheless "they cannot speak as we do, that is they cannot show that they are thinking..." (CSM I, p. 140). He holds that if the fact that animals excel in some tasks more than humans would indicate that these tasks require intelligence, then animals should be more intelligent than humans and excel in all tasks. Since this clearly is not the case, Descartes concludes that those tasks do not require intelligence:

> ...so what [animals] do better does not prove that they have any intelligence, for if it did they would have more intelligence than any of us and would excel us in everything. It proves rather that they have no intelligence at all, and that it is nature, which acts in them according to the dispositions of their organs. In a same way a clock, consisting only of wheels and springs, can count the hours and measure time more accurately than we can with all our wisdom. (CSM I, p. 141)

Here Descartes stresses that 'nature' allows creatures to excel in some things while intelligence allows us to excel in other things. One of these latter things is using language. In a similar manner he stresses in *Letters to More* that "speech is the only certain *sign* of thought" (CSMK III, p. 366, emphasis added) and that "speech, ... alone shows the thought hidden in the body" (CSMK III, p. 374). In the *Letter to the Marquess of Newcastle* Descartes says he believes "that the reason that animals do

not speak as we do is not that they lack the organs but that they have no thoughts" (CSMK III, p. 303). He reiterates this point throughout the letter and objects to the proposition that "…[animals have] some thought such as we experience in ourselves, but of a very much less perfect kind" (CSMK III, p. 304). His reply is unambiguous: "…if [animals] had thought as we do, they would have to have immortal souls like us. This is unlikely, because *there is no reason* to believe it of some animals without believing it of all, and many of them such as oysters and sponges are too imperfect for this to be credible (Ibid., emphasis added). These passages confirm again that Chomsky (1975a) is incorrect to suggest that Descartes thought animals could not develop a language of the human type because they have limited intelligence. And it also shows that Descartes would have disagreed with this Chomskyan reconceptualization of his view: "Descartes would not have regarded the acquisition of language (or other cognitive faculties) as problematic and never proposed they offered a challenge to the mechanical philosophy" (Chomsky, 2010b, p. 104). This claim stands in contrast to the passages just cited and I believe we should consider Descartes to be the ultimate authority on the matter. Descartes writing indicates that language does not depend on a special organ (like a language faculty) but that language is the outward expression of thought (Miel, 1969; Cottingham, 1992. 1997; Schmaltz, 1997; Erion, 2001; Grush, 2003; Williams, 2005).

Descartes observed that (at least some) animals do have bodily organs that are sufficiently similar to the bodily organs involved in human speech. Furthermore, he reports on cases in which animals have been taught to produce some words of human language. This indicates that having the organs dedicated to language production is

not sufficient for language use. What is required is a rational mind, and, according to Descartes, animals lack such a mind (Gunderson, 1964; Miel, 1969; MacDonald Ross, 1988; Cottingham, 1978, 1992; Morris, 2000; Gaukroger, 2000; Erion, 2001; Alanen, 2003; Williams, 2005; Schmaltz, 2006; Hoffman, 2007; Hatfield, 2008b). But clearly humans have language because they have rational minds (not the other way around).

A further point in support of my argument is that for Descartes minds are indivisible: "we cannot understand a mind except as being indivisible. For we cannot conceive of half a mind" (CSM II, p. 9). In *Rules for the Direction of the Mind* Descartes states that our knowledge depends only on a purely spiritual power which is "*one single power*...it is *one and the same power*. ...According to its different functions... the same power is called either pure intellect or imagination or memory or sense perception" (CSM I, p.42, emphasis added). The same fact is explained from a different perspective in the *Sixth Meditation*:

> ...minds are utterly indivisible. For when I consider the mind, or myself in so far as I am merely a thinking thing, I am unable to distinguish any parts within myself... As for the faculties of willing, of understanding, of sensory perception and so on, these cannot be termed parts of the mind, since it is one and the same mind that wills, and understands and has sensory perception. (CSM II, p. 59)

It is important to note that Descartes does not mention language as a possible faculty here. In fact the two intellectual faculties he mentions (willing and understanding) both 'use' language. While it is possible to claim that a domain-specific language faculty could be contained in 'and so on', it appears more plausible

to suggest that our linguistic abilities depend on the mind's essence of rational thought.

Descartes stresses repeatedly that the essence of mind (soul) is thought (Malcolm, 1968; Kenny, 1972; McRae, 1972; Adams, 1975; Marles, 1978; Bracken, 1984; Sorell, 1987; Baker& Morris, 1996; Flage & Bonnen, 1999; Schmaltz, 2002, 2008; Gaukroger, 2002; Hoffman, 2002, 2007; Alanen, 2003; Williams, 2005). In *Discourse on Method* he writes that his mind (soul) is "a substance whose whole essence or nature is simply to think" (CSM I, p. 127). In the *Conversation with Burman* Descartes maintains that "the mind can never be without thought; it can of course be without this or that thought but it cannot be without *some* thought. In the same way, the body cannot, even for a moment, be without extension" (CSMK III, p. 336, original emphasis). The same thought is expressed again in a *Letter for Arnauld:* "thought constitutes [the mind's] essence... Thought is not conceived as an attribute which can be present or absent" (CSMK III, p. 355). And in a *Letter to Hyperaspistes* Descartes does not endorse an internal structure of the mind but the mind's ability to *acquire* new properties, some of which could be occurent ideas: " ...even though the mind is indivisible, it is none the less capable of acquiring various properties" (CSMK III, p. 196).

Finally, in *Rules for the Direction of the Mind* Descartes insists that all knowledge depends on the intellect: "... nothing can be known prior to the intellect, since knowledge of everything else depends on the intellect" (CSM I, p. 30), and the only other "instruments of knowledge... [are] imagination and sense-perception" (Ibid.). Later he mentions a fourth faculty: "while it is the intellect alone that is

capable of knowledge it can be helped or hindered by three other faculties, *viz*.

imagination, sense-perception and memory" (CSM I, p. 32). If Descartes believed

that our knowledge depends also on a domain specific language faculty (Cowie,

1999) it is curious that he does not mention such a faculty in his enumerations. Since

language is 'used' by at least three of these faculties one would be led to the

assumption that Descartes believed that language is domain-general not domain-

specific.

Chomsky observes that Descartes' view regarding the indivisibility of the

mind is quite different from his own proposal that language relies on a domain-

specific organ. He could acknowledge the difference and suggest that, nevertheless,

there are important parallels between his own and Descartes' view. But Chomsky

chooses a different path:

> For Descartes the mind is not part of the biological world. Furthermore he
> appears to regards the mind as uniform and undifferentiated: 'there is
> within us one soul and this soul has not in itself any diversity of parts';
> the mind is entirely indivisible' [HR, p. 353; p. 196] One might then
> argue that we are not studying Descartes' problem when we consider the
> human mind as a specific biological system, one with components and
> elements of a varied kind, to be explored as we would study any other
> aspect of the physical world. This conclusion holds, however *only* if we
> regard Descartes as an irrational dogmatist, that is, as putting forth
> doctrines that define the domain of inquiry, rather than as arguing for
> principles that he believed he had established within an inquiry more
> broadly construed. That seems to me a questionable move. (Chomsky,
> 1980, pp. 30-31, emphasis added)

The use of 'only' suggests that, if we attribute to Descartes the view he

actually held we regard him as an irrational dogmatist. Chomsky provides no

argument supporting his conclusion that Descartes' actual view would be irrational dogmatism. The most charitable interpretation I can imagine is that Chomsky thinks, in the light of the evidence we have now, Descartes would be an irrational dogmatist if he would continue to hold a view that was based on the evidence that was available 350 years ago. Descartes might indeed have held a different view, had lived in the 20th century. He might have agreed with Chomsky that "there is no reason to accept the traditional doctrine, as expressed by Descartes, that all human reason is a 'universal instrument which can serve for all contingencies" (Chomsky, 1980, p. 251). But Descartes might also have taken a 'connectionist turn' and defended the general-purpose view of intelligence that Chomsky opposes. As some of the passages I have cited indicate, this is not entirely implausible. Another possibility is that a 'modern-day' Descartes would adopt a view that is similar in spirit to that defended by Katz (1981, 1998). Or he could maintain dualism and endorse a combination of connectionist empiricism and rational realism proposed by Katz & Postal (1991). One could make arguments for every one of these possibilities.

Yet, this kind of speculation does not appear to be Cartesian in spirit. We do know that when Descartes was still able to comment on such matters he did not cherish the attempts of others to 'improve' his writings: "I find this learned doctor's treatment of my writings and his efforts at interpreting (or rather, falsifying) them much more annoying than the most bitter attacks which others have made upon them" (CSM II, p. 308). For this reason I suggest we should refrain from speculation and focus on what Descartes actually wrote.

In conclusion, while some interpreters (Chomsky, 1966, 1975b, 1980; Cowie,

1999; McGilvray, 2002; Matthews, 2006) have suggested that Descartes might endorse a domain-specific language faculty, we find little textual evidence supporting this claim. It appears more plausible that Descartes held that humans can acquire and use language because they have rational minds. Our rational minds underwrite language acquisition and use and allow for other forms of intelligent behaviour. Some commentators (Erion, 2001; Alanen, 2003) note that the Cartesian test for intelligence with its focus on language and intelligent action is more demanding than the Turing test, which focuses on language alone. For Descartes language is an essential characteristic of humans (species specific). But it is only an indicator of a rational mind not its foundation. I will turn now to Descartes' view of ideas and inquire whether or not this view commits him to a Chomskyan theory of language acquisition.

2.3. Language Acquisition

Before discussing details of Descartes' doctrine of ideas I would like to point out that one focus of this chapter is language acquisition. An important point in support of Chomsky's claims about language acquisition is his poverty of the stimulus argument. Broadly speaking from the premises (i) that the input received by the language-learning child is insufficient to explain her output if this output depends on data-driven general-purpose learning mechanisms alone and (ii) that, nevertheless, virtually all children acquire language competence with seeming ease, Chomsky draws the conclusion that a domain-specific innate structure (the LAD) must exist.

Allegedly, this LAD allows children to infer rules of grammar from an insufficient inductive database and to become linguistically competent speakers. The LAD is a physical structure in the biological brain, and language acquisition is a purely mechanical process. Chomsky always held that there are important parallels between his own view and Descartes' doctrine of innate ideas. I shall inquire now whether Descartes' writings provide supporting evidence for Chomsky's claims.

For the following discussion it is important to stress that Chomsky is concerned with the language acquired by young children during the first years of their lives. Allegedly these children have an insufficient language input to account for the output they produce, and they require some form of innate structure that helps them in the task. Chomsky (1966, 1975a, b, 1980, 1985, 2009b, 2010a) and several of his supporters (e.g., McGilvray 2002; 2009, Matthew, 2005) and critics (e.g., Cowie, 1999) hold that Descartes was committed to a similar view. They invariably cite Descartes' writings on innate ideas in support of this claim. However, very little attention is paid to what Descartes wrote specifically on the acquisition of language and the difference between the ideas we acquire in our childhood and later in life.

I will first cite some of these passages and then discuss how they may affect an interpretation of Descartes' view on language acquisition. One of the most explicit comments about the first acquisition of language can be found in *Search for Truth,* where Eudoxus compares the acquisition of scientific knowledge with 'simple forms of knowledge': "I should like you to notice how the sciences differ from those simple forms of knowledge which can be acquired without any process of reasoning, such as *languages*, history, geography and in general *any subject which rests on experience*

alone" (CSM II, p. 403, emphasis added). This passage seems to confirm the view that language acquisition could be indeed a purely mechanical process. However, it is not clear from the context, whether Descartes refers here to first- or second-language acquisition. Thus, it would be premature to argue that this passage commits him to the view Chomsky (2010b) ascribes to him. However, there are several other passages that might indicates that the traditional rationalist-nativist interpretation of Descartes is problematic as well.

In a *Letter to the Marquess of Newcastle* Descartes stresses the fundamental difference in language capacities between humans and animals and makes the point that "even deaf-mutes *invent special signs* to express their thoughts" (CSMK, III, p.303, emphasis added), and in *Discourse on Method* the same claim is made " ...men born deaf and dumb... *normally invent their own signs* to make themselves understood by those who, being regularly in their company, have the time to learn their language" (CSM, I, p. 140, emphasis added). In these passages Descartes' main emphasis seems to be on the communicative function of language. If, as Chomsky often claims, language would be mainly a tool to express one's thought, the deaf-mute would probably not bother to invent language-related signs to communicate.

Furthermore, we can easily find evidence suggesting that Descartes believes that, for language acquisition, people rely on learning instead of a process of 'automated' parameter setting of an innate language organ. In a *Letter to Mersenne* he asserts that "[t]here are only two things *to learn* in any language: the meaning of words and grammar" (CSMK, III, p. 10, emphasis added) and in a *Letter to Chanut* he explains in some detail how we acquire the meaning of words:

> ... when we learn a language, we connect the letters or the pronunciation of certain words, which are material things, with their meaning, which are thoughts, so that when we later hear the same words, we conceive the same things, and when we conceive the same things, we remember the same words. (CSMK, III, p. 307)

Here Descartes discusses the essential connection between sense perception (objects and sound combinations) and mental concepts for language acquisition. Furthermore, Descartes repeatedly states that language arbitrarily connects sounds to objects and that we rely on memory when we recall words we have heard. In the *Conversation with Burman* he explains:

> When for example on hearing that the word "K-I-N-G" signifies supreme power, I commit this to my memory and then subsequently recall the meaning by means of my memory, it must be intellectual memory that makes this possible. For there is no relationship between the four letters (K-I-N-G), which would enable me to derive the meaning from the letters. It is intellectual memory that enables me to recall what the letters stand for. (CSMK III, pp. 336-7)

This account is quite compatible with *any* current account of language learning. On the one hand it does not *require* any faculty more domain-specific than 'intellectual memory'. On the other hand it would be compatible with the domain-specific LAD hypothesis.

Concerning the content of ideas, Descartes repeatedly points out that the ideas we acquire in our childhood are often "confused thoughts" (CSMK, III, p. 308), based on "false preconceptions" (CSMK, III, p. 233). Usually we learn only much later

about the true nature of things and acquire what Descartes calls 'clear and distinct ideas'. This is again, very different in spirit from the 'instantaneous' acquisition of grammar and semantic concepts in Chomsky's account. One example for the 'Cartesian' progression from confused to clear and distinct ideas can be found in the discussion of our different ideas of the sun:

> ...there are two different ideas of the sun which I find within me. One of them which is acquired as it were from the senses and which is a prime example of an idea which I reckon to come from an external source, makes the sun appear very small. The other idea is based on astronomical reasoning, that is, it is derived from certain notions that are innate in me (or else it is constructed by me in some other way), and this idea shows the sun to be several times larger than the earth. (CSM, II, p. 27)

For my purpose it is important that Descartes calls the idea that is based on astronomical reasoning 'innate'. Yet usually we only 'access' this idea long after we have acquired language. It is quite evident that Descartes does not deny that we have the first idea of the sun. Moreover, children probably have only this idea. Of course, that does not prevent them from using the word 'sun' in appropriate circumstances, creatively and stimulus-free. Even young children understand that the word 'sun' refers to a certain object, and they do not need to see the sun every time they talk about it. For everyday communication either idea of the sun will do just fine. Only when we consider questions about the 'true nature' of the sun, do we need the innate idea. One could, of course, argue that Descartes would be committed to the view that we are unable to speak about the sun unless we have a true understanding of 'sun', unless we access the innate idea. But this seems a far stretch from what Descartes *is*

saying. In fact his attempts to distinguish true science from everyday (mis)conceptions indicate that he is quite aware of the fact that language acquisition and use do not rely on a deep understanding of the true nature of things.

A similar thought is expressed in *Search for Truth* when Descartes refers to the learning in early childhood: "...he came into the world in ignorance and ... the knowledge he had as a child was based solely on the weak foundation of the senses and the authority of his teachers" (CSM II, p. 400). Undoubtedly, part of the knowledge we acquire in early childhood is knowledge of language. And while Descartes makes it clear that science could not be placed on the shaky foundation of empirical knowledge, he does not seem to have similar concerns about language.

Furthermore, language seems never to be the problem when Descartes discusses the unreliability of the senses. For example, when he deals with the wax example in the *Second Meditation,* Descartes challenges us to doubt virtually everything about the wax (other than that it is an extended substance). Yet, not once does he question whether or not we are able to attach the right label to 'wax'. He never questions that when different people use the word 'wax' they refer to the same perceivable object. Here as elsewhere, his discussion of the different characteristics of the wax and our inability to decide, based on sense experience alone, what its true nature is presupposes that we use language appropriately, creatively and stimulus free. What is at issue is whether wax is really what the senses (vision, touch, imagination) seem to reveal to us or whether the true nature of wax can only be grasped by the rational mind: "... *the nature* of the piece of wax is in no way revealed by my imagination but perceived by the mind alone" (CSM, II, p. 21, emphasis

added).

For Descartes this perception by the mind can either be 'imperfect and confused' or 'clear and distinct'. Now what makes the important difference to 'true' knowledge is "how carefully I concentrate" (put my mind to use) not how/if I use language. In fact Descartes notes that "although I am thinking about these matters within myself, silently and without speaking, nonetheless *the actual words bring me up short, and I am almost tricked by ordinary ways of talking*" (CSM, II, p. 21, emphasis added). This passage seems to indicate that while words (language) are a tool of the mind, this tool is by no means perfect, and it can not only help but also hinder the proper working of the mind.

There is a similar example in *Principles of Philosophy*. Here Descartes explains that our use of language (words) can cause us to commit errors because the words we use do not always correspond to the concepts they signify:

> ... because of the use of language we tie all our concepts to the words used to express them; and when we store the concepts in our memory we always simultaneously store the corresponding words. Later on we find the words easier to recall than the things and because of this it is very seldom that our concept of a thing is so distinct that we can totally separate it from our concept of the words involved. The thoughts of almost all people are more concerned with words than with things; and as a result people very often give their assent to words they do not understand, thinking they understood them, or that they got them from others who did understand them correctly. (CSM I, p. 220-21)

Descartes notes that the words of our language are imperfect tools (Ariew et al., 2003) for the expression of our thoughts, *not* integral parts of our innate concepts. The fact that we need to store our concepts simultaneously with words could indicate that there is no innate connection between the two. Furthermore, much of language

acquisition is seemingly an empirical affair. The words are learned either by association to objects or by instruction from others. And, of course, both methods are prone to error. Finally, Descartes seems to allow that experience can change not only our way of thinking but also our use of language:

> ... nature seems to have joined every movement of the [pineal] gland to certain of our thoughts from the beginning of our life, yet we may join them to others through habit. Experience shows this in the case of language. Words produce in the gland movements which are ordained by nature to represent to the soul only the sounds of their syllable when they are written, because *we have acquired* the habit of thinking of this meaning when we hear them spoken or see them written. (CSM I, p. 348, emphasis added)

This passage, which uses specifically the example of language, contains two problems for the Chomskian nativist. First, Descartes claims that we have *acquired the habit* of connecting certain words with certain meanings. Second, Descartes allows that habit can change the 'natural' connections between words and meanings. This seems to be at odds with Chomsky's claim that language acquisition is a mechanical triggering of innate concepts and to fit better with an associationist view of language acquisition.

So far I have explored a very unusual hypothesis regarding Descartes' view about language acquisition. He states that language "can be acquired without any process of reasoning... [based] on experience alone" (CSM, II, p. 403), that we learn language by connecting words with their meanings and remembering later upon encountering words which things they signify and vice versa (CSMK, III, p. 307), at a time when our thoughts are 'confused' and based of 'misconceptions'. Not

68

surprisingly, the language we acquire under such circumstances is not a perfect tool for the correct expression of our thoughts. But while we can (and should) acquire a new way of thinking Descartes does not suggest that our language needs to be changed. Having established these points about language acquisition and use I need to inquire whether or not this reading of Descartes is compatible with his views about ideas in general and innate ideas in particular. To this task I shall turn next.

2.4. Cartesian Ideas

It is well established that Descartes uses the term 'idea' differently from contemporary usage and that he does not always mean the same when using 'idea' (e.g., McRae, 1965, 1972; Kenny, 1967, 1968; Doney, 1967; Gewirth, 1967; Ashworth, 1972, 1975; Ree, 1974; Adams, 1975; Danto, 1978; Tlumak, 1978; Markova, 1982; Costa, 1983; van de Pitte, 1985; Bolton, 1986; Chappell, 1986; Cottingham, 1978, 1986, 1992; Normore, 1986; Jolley, 1988; Wells, 1990; Gaukroger, 1992; McGraw, 1992; Nelson, 1997; Schmaltz, 1997; Vinci, 1998; Flage & Bonnen, 1999; Behan, 2000; Boyle, 2000; Sepper, 1996; Schmaltz, 2002; Alanen, 2003; Ariew et al., 2003; Clarke, 2003; Hatfield, 2003; Williams, 2005; Nadler, 2006; Rozemond, 2006; Hatfield, 2008a; Vinci, 2008).

To avoid misinterpretation it is important to pay close attention to both the potentially different meanings of Descartes' term 'idea' and the specific context in which he uses 'idea' in a given passage. In *Preface of the Meditations* Descartes cautions that the word 'idea' is ambiguous because "the term idea… may be taken

either materially for an act of the understanding, … or objectively, for the thing represented by that act" (CSM II, p.7). Throughout his writings Descartes switches frequently between these two meanings. For example in *Meditation Three* he first seems to insist that ideas are objects of the mind: "Some of my thoughts are as it were images of things, and it is *only in these cases* that the term 'idea' is strictly appropriate - for example when I think of a man, or a chimera, or the sky, or an angel, or God" (CSM II, p. 25, emphasis added). Yet, when he continues "... if I considered just the ideas themselves, simply as modes of my thought" (CSM II, p. 26), he has turned to operations of the mind. Later, when he distinguishes among adventitious, fictitious, and innate ideas, he seems to talk about both objects and operations of the mind. Here he defines adventitious ideas as ideas derived from our experience of the world ('things existing outside me" CSM II, p. 26). Fictitious ideas are defined as ideas that are invented by the imagination. By contrast, innate ideas are not dependent on perceptions but "derive simply from my own nature" (CSM II, p. 26) they are "true ideas which are innate in me", CSM II, p. 47). In *Meditation Five* Descartes seems to refer to ideas as objects in the mind again when he writes: "… whenever I do choose to think of the first and supreme being, and bring forth the idea of God from the treasure house of the mind…" (CSM II, p. 46) and when he describes the idea of god as "an image of a true and immutable nature" (CSM II, p. 47).

In addition to the ambiguity of the word 'idea' there seem to be some inconsistencies in Descartes' application of the three sub-categories: adventitious, fictitious, and innate ideas. On the one hand we have the frequently cited passages in *Comments on a Certain Broadsheet* where Descartes seemingly asserts that all ideas

are innate:

> ...there is nothing in our ideas which is not innate to the mind or the faculty of thinking... the very ideas of motions themselves and of the figures are innate in us. The ideas of pain, colors, sound, and the like must be all the more innate if, on the occasion of certain corporeal motions, our mind is to be capable of representing them to itself, for there is no similarity between these ideas and corporeal motions. (CSM I, p.304)

This passage seems to suggest an extreme innatism, and, accordingly, it has been frequently marshaled in support of such a reading of Descartes' commitment (e.g., Cowie, 1999; Chomsky, 1968, 1975a, b; Katz, 1971; Bracken, 1984; Flage & Bonnen, 1999; McGilvray, 2002).

But others (e.g., Windelband, 1958; Stich, 1975; Adams, 1975; Nelson, 1997; Schmaltz, 1997) suggest that a more careful reading of the context reveals that Descartes' main goal here was to refute a scholastic account of perception. Descartes rejected the scholastic doctrine that sensations resemble qualities of bodies and are literally transmitted from the object of perception to the mind (Danto, 1978; Sorell, 1987; McGraw, 1992; Schmaltz, 1997; Vinci, 1998; Flage & Bonnen, 1999; Wolf-Divine, 2000; Rozemond, 2006). On this interpretation what is innate is not the content of sensory ideas but the faculty of sense perception (Adams, 1975; Schmaltz, 1997; Nelson, 1997; Flage & Bonnen, 1999; Ariew et al., 2003; Rozemond, 2006). This interpretation appears to be plausible for two reasons.

First, Descartes repeatedly establishes the distinctions among three types of ideas (adventitious, fictitious, and innate), and he provides numerous examples for adventitious ideas that are caused by experience. The Cartesian classification system

for ideas occurs in the *Meditations* (CSM II, p. 26, 47) the *First Set of Replies* (CSM II, p. 76) and in several letters (e.g., to Mersenne: CSMK III, p. 183; to Burman: CSMK III, p. 347; to Clerselier: CSMK III, p.376). In addition specific reference to innate ideas can be found in *Comments on a Certain Broadsheet* (CSM I, p. 309; equating innate ideas with a natural power to know God); the *Third Set of Replies* (CSM II, p. 132; equating innate ideas with a faculty to summon up ideas), the *Fifth Set of Replies* (CSM II, p. 258; discussing the innate idea of God and the self and p. 262, discussing innate ideas of geometrical objects), a *Letter to Hyperaspistes* (CSM III, p. 190; referring to the innate ideas of God, the self and self-evident truths in an infant's mind), in *Conversation with Burman* (CSM III, p. 336; reference to the innate idea of God).

The foregoing shows that (i) Descartes develops and employs a classification system of different kinds of ideas and (ii) that innate ideas are seen as either a subset of the entire set of ideas or a (intellectual) faculty of the mind. There seem to be occasional inconsistencies (see below) in Descartes' application of the idea-classification system. But we find little support for a literal reading of the Broadsheet passage (CSM I, p.304) if we take 'idea' there to refer to content. Plausibly what are innate are dispositions to form occurrent ideas with a certain content, but the content itself is not innate. We find several passages where Descartes specifically refers to ideas that get their content from the 'outside' world, which means that this content cannot be innate. In *Meditation Six* he allows that at least some ideas are transmitted to us through sense experience of material objects. He claims that imagination is "nothing else but an application of the cognitive faculty to a body which is intimately

present to it, and which therefore exists" (CSM II, p. 50). While it is not clear yet whether or not this body exists independently and outside of the mind, Descartes continues, "...when [the mind] imagines it turns toward the body and looks at something in the body which conforms to an idea understood by the mind or perceived by the senses" (CSM II, p. 51). Throughout *Meditation Six* Descartes inquires whether or not the senses provide us with reliable perceptions of external objects. Yet, he does not deny that the senses perceive *something*.

Similarly, in *Fifth Set of Replies* Descartes asserts that he "…showed that [ideas of material things] often come to us from bodies, and that it is this that enables us to prove the existence of bodies" (CSM II, p.253). In *Discourse on Method* Descartes explains, "light, sounds, odors, tastes, heat, and all the other qualities of external objects can *implant* various ideas through the medium of the senses" (CSM I, p. 139, emphasis added). This indicates that the content of these ideas has not been innately there but is derived from the external world. In a *Letter to Clerselier* Descartes again holds that at least some ideas are not innate: "...I was also obliged to distinguish the ideas which are born with us from those which come from elsewhere, or are made by us..." (CSMK III, p. 376). Finally, in *Conversation with Burman* Descartes gives a specific example of ideas that do get their content from external sources: " there are also some [ideas] which are adventitious, for example the idea of Leiden or Alkmar" (CSMK III, p. 347). Taken together these passages make it implausible to suggest that Descartes was committed to the belief that the content of all ideas is innate.

For my inquiry it is not essential whether or not Descartes can eventually

provide a coherent argument for the claim that the senses can provide us with knowledge about external objects. My aim is far more modest. I want only to refute the assertion that Descartes was committed to the kind of innatism suggested by Cowie (1999). Towards the end of *Meditation Six* Descartes has eliminated the hyperbolic doubt about external bodies and reminded us that we often can use several of our senses for cross-reference: "I can always almost make use of more than one sense to investigate the same thing; and in addition I can use both, my memory which connects present experiences with preceding ones, and my intellect, which has by now examined all the causes of error" (CSM II, p. 61). Here Descartes overcomes the doubt regarding the external world not based on innate ideas but based on the integration of information provided by our senses, memory and intellect: "when I distinctly see where things come from and where and when they come to me...I ought not to have even the slightest doubt of their reality if, after calling upon all senses as well as my memory and my intellect...I receive no conflicting report from any of these sources" (CSM II, p. 62). These passages challenge the claim that for Descartes all *content* of ideas is innate. They indicate that at least some ideas are adventitious as defined by him earlier.

Further, in some instances in which Descartes seemingly speaks about innate content of ideas, alternative interpretations are possible. In *Principles of Philosophy* he asserts that the mind "finds within itself ideas of many things" (CSM I, p. 197). However, it is left open how these ideas got there. Descartes continues, "so long as [the mind] merely contemplates these ideas and does not affirm or deny the existence outside itself anything resembling them, it cannot be mistaken" (Ibid.). This allows

for the possibility that the content of at least some of these ideas originates from external sources. Otherwise we would never be correct to affirm the existence of outside objects based on the ideas we have of them.

By the end of the *Sixth Meditations* Descartes has removed the complete doubt regarding the external world. Of course, accepting that the outside world exists does not provide an argument against skepticism about individual external objects. Even though we can be assured that some objects of our sensory experience exist, we still can be mistaken about specific objects. Therefore we need to be careful before we *affirm or deny* the existence of objects that resemble our sensory experiences. But Descartes allows for the possibility that our ideas at least sometimes provide evidence for the existence of external objects. This casts some doubt on the belief that he was committed to an extreme innatism regarding the content of ideas.

As the previous passages have shown, Descartes would face a dilemma if he were committed to the extreme innatism attributed to him by Cowie (1999). On the one hand it would be difficult to understand why he would allow for the possibility that the content of at least some ideas originates from outside sources. On the other hand it would be difficult to maintain that we can know anything about the external world, if the content of all ideas originates from within us. Seemingly, Descartes resolves this dilemma in a *Letter to Hyperaspistes,* where he clearly asserts that not all ideas are innate: "I proved the existence of material things not from the fact that we have ideas of them but from the fact that *these ideas come to us in such a way as to make us aware that they are not produced by ourselves but come from elsewhere*" (CSMK III, p. 193, emphasis added). For Descartes' ontology and epistemology it is

important that *not* all content of ideas originates in our minds. This allows him to escape solipsism and to rely on a causal theory of (the content of) ideas when proving the existence of the physical world. But while this task is of great importance for Descartes' overall philosophical project, it has little to do with his views regarding language. Therefore we shall return to the Cartesian classification system of ideas.

Descartes makes yet another distinction regarding ideas. He distinguishes between ideas of the mind and ideas of the imagination when he writes in a *Letter to Mersenne*: "whatever we conceive without an image is an idea of the pure mind and whatever we conceive with an image is an idea of the imagination" (CSMK III, p.186). Since for Descartes the imagination is associated with the body, it is reasonable to assume that (at least some) ideas of the imagination would derive from external sources and not be innate (Vinci, 2008). I propose that we have enough evidence to reject the claim that Descartes held that 'everything is innate'. Instead, there is a complex interaction between the resources of the mind (ideas as operations of the mind) and external and mental objects (content of ideas). In the next section I will look at 'innate ideas' as used by Descartes and inquire whether or not Descartes' doctrine of 'innate ideas' would support the view of language acquisition that Chomsky calls Cartesian.

2.4.1. Innate Ideas

In order to understand what Descartes' innatism commits him to, we need to determine what is innate on his account: operations of the mind, objects of the mind

or both. Apparently Descartes answers this question differently in different places. On one hand he seems to refer to the content of ideas when he talks about innate ideas. In *Meditation Five* he suggests that innate ideas appear to be remembered: "on first discovering [innate ideas] it seems that I am not so much learning something new as remembering *what* I knew before" (CSM II, p. 44, emphasis added). Clearly, Descartes speaks here not about operations of the mind but about (specific) content. This is reminiscent of Plato's theory of recollection, and a similar point is made more explicitly in a *Letter to Voetius*:

> ...notice that all things whose knowledge is said to be naturally implanted in us are not for that reason expressly known by us; they are merely such that we come to know them by the power of our native intelligence, without any sensory experience. All geometrical truths are of that sort... Socrates asks the slave boy about elements of geometry and thereby makes the boy able to dig out certain truth from his own mind, which he had not previously recognized were there.... Our knowledge of God is of this sort... (CSMK III, p. 222-23)

Here Descartes speaks again about some specific content (e.g., knowledge of geometrical truth or of God) that has been naturally implanted in us. Yet, he also refers to the power of our native intelligence. And, in other contexts, he refers clearly to operations of the mind; for example in his *Reply to the Third Set of Objections* Descartes explains that by innate "…we simply mean that we have within ourselves the faculty of summoning up the idea" (CSM II, p. 132). Similarly, in *Reply to the Seventh Set of Objections* he states that: "there is present in us a power to know…" (CSM, II, p.309).

These thoughts have been taken to foreshadow Chomsky's postulation of an

innate domain-specific language acquisition faculty (e.g., Chomsky, 1968, 1975b; Cowie, 1999; McGilvray, 2002, 2009; Matthews, 2005). These authors claim that Descartes commitment to a faculty that is not only innate but also domain specific is supported by an analogy he provides in the *Comments on a Certain Broadsheet*. Here he likens innate ideas to heritable diseases and points out that "it is not so much that the babies of such families suffer from these diseases in their mother's womb; but simply that they are born with a certain 'faculty' or tendency to contract them" (CSM I, p. 304). Descartes does not compare innate ideas with our (general) tendency to contract diseases but with an innate predisposition to contract *certain* diseases. Cowie (1999) suggests that Descartes holds that by analogy it would follow that the mind possesses special structures (faculties) that are dedicated to the acquisition of particular sorts of ideas. However, this would contradict Descartes' repeated assertion that he believes the mind is indivisible. Therefore, we need to pay close attention to how the term 'idea' is used. In the same passage Descartes writes:

> I have never...taken the view that the mind requires innate ideas which are something *distinct* from its own *faculty of thinking*...I did however, observe that there were certain thoughts within me which neither came to me from external objects nor were determined by my will but which came solely from the *power of thinking* within me; so I applied the term 'innate' to the ideas or notions which are the forms of these thoughts... (CSM I, p. 303, emphasis added)

Pace Cowie it appears that here, when discussing innate ideas, Descartes refers to the (general) faculty of thinking not to a *specialized* faculty of language. The faculty of thinking undoubtedly includes language, but I will argue below that it cannot be reduced to language. If this correct, then the above passage cannot be read

as supporting evidence for a domain-specific language faculty theory. This is a point that Chomsky seems to accept when he admits that Descartes held the doctrine "that human reason is a 'universal instrument, which can serve for all contingencies'" (Chomsky, 1980, p. 251). I suggest that the evidence for domain-specificity in the *Broadsheet* passage is much weaker than Cowie suggests.

A particularly relevant example for my inquiry can be found in a *Letter to Mersenne*:

> I use the word 'idea' to mean everything which can occur in our thought, and I distinguish three kinds. Some are adventitious, such as the idea we commonly have of the sun; others are constructed or made up, in which class we can put the idea which the astronomers construct of the sun by their reasoning; and others are innate, such as the idea of God, mind, body, triangle, and in general all those which represent true immutable and eternal essences. (CSMK III, p. 183)

This passage is important because it shows that in this letter Descartes considered the idea of the sun 'as constructed by the astronomers', not as innate but as 'made-up'. While it is possible to hold that 'made-up' ideas are essentially innate, two conditions need to be met. First, the operations of the mind that allow us to 'make-up' these ideas need to be innate and second, the content of the 'building blocks' of these made up ideas (e.g., the simple ideas contained in them) needs to be innate. Descartes discusses these points in his *Conversation with Burman*. Using the example of our idea of the trinity he writes:

> ...even though the idea of the Trinity is not innate in us to the extent of giving us an *express representation* of the Trinity, none the less the

elements and rudiments of the idea are innate in us, as we have an innate idea of God, the number 3, and so on. It is from these rudiments, supplemented by revelation from the Scriptures, that we easily form the full idea of the mystery of the Trinity. This is how the conception we have of it is formed. (CSMK III, p. 347, emphasis added)

This passage suggests that 'rudiments' or simple ideas are innate on Descartes' view. If this is the case, then we have here an important difference between Descartes and Chomsky's views regarding the contribution of innate ideas/knowledge to language acquisition. As we have previously seen, when Descartes speaks about innate ideas, he refers either to operations of the mind or to (semantic) content of ideas. He says virtually nothing about the syntactic rules of language and how they are acquired. And the 'innate (semantic) content' on Descartes' view (simple elements or rudiments) is something that (with the exception of 'God' or perfect triangles) *could* have been acquired from sense experience (or available input). By contrast, for Chomsky what is innate is not only semantic content but also syntactical or structural principles that could not have been acquired from the available input. These principles only can be acquired if we postulate innate domain-specific structures. These structures allow for the acquisition of complex semantic content and syntactical principles. This suggests that the innate resources make very different contributions to language acquisition on Chomsky's and Descartes' views.

At this point I would like to return to the *Comments on a certain Broadsheet* passage. Descartes writes: "The ideas of pain, colors, sound, and the like must be all the more innate if, on the occasion of certain corporeal motions, our mind is to be capable of representing them to itself, for there is no similarity between these ideas

80

and corporeal motions" (CSM I, p.304). It appears that he does not mean that the 'semantic content' of these ideas is innate. Rather, he seems to suggest that our ability to perceive colours *as* colours, sounds *as* sounds etc. is innate. This reading is supported by the text. In *Rules for the Direction of the Mind, Rule 14* Descartes writes:

> ...if someone is blind from birth, we should not expect to be able by force of argument to get him to have *true ideas* of colours just like the ones we have, *derived* as they are *from the senses*. But if someone at some time has seen the primary colours, though not the secondary or mixed colours, then by means of a deduction of sorts, it is possible for him to form images even of those he has not seen *in virtue of their similarity to those he has seen.* (CSM I, p. 56-57, emphasis added)

Several points are important here. In the first part of the argument (which reminds us of Frank Jackson's (1982) knowledge argument), Descartes emphasizes the difference between someone who has only heard everything about a colour and someone who has actually seen this colour. This would indicate that the innate idea of colour mentioned in the *Comments on a certain Broadsheet* passage cannot be a complete, true idea of colour because in that case the blind person should have (at least potentially) the *same* idea as a seeing person. However, like Jackson, Descartes insists that there must be a difference between the two ideas, and he explains this difference with sense experience or lack thereof. In the second part of the argument (which reminds us of David Hume's (1739) missing shade of blue argument) Descartes suggests that someone who has seen *some* colours could deduce images of colours he has not seen before. This process seems to rely on at least two

components: the rational mind with its faculty of thought (and memory) *and* the actual sense perception. Descartes seemingly does not suggest that the missing colors are already in the mind and could be accessed by 'activating' a latent or potential image of them. Instead, the mind seems to need at least some previous sense experience of colour. Finally, it is important to note that Descartes specifically requires that the colours the person has seen are the primary colours, the 'colour-simples'. From these building blocks he might be able to 'assemble' secondary or mixed colours.

Descartes makes a similar argument in a *Letter to Hyperaspistes* where he asserts that even if a man born blind would have "ideas exactly like our ideas of colours: he still cannot know that they are like ours, or that they are called ideas of colours, because he does not know what ours are like" (CSMK III, pp. 195-6). While this statement seems to imply some innatism (the blind person could have ideas of colours just like ours), Descartes immediately refutes this interpretation. By saying the blind man 'could not know' that his ideas are like ours because he 'does not know what ours are like' Descartes denies two things. First, if the content of colour ideas would be supplied innately, then there would be no reason to deny that the blind man knows that his ideas are the same as ours. Second, by saying that the blind man does not know that his ideas of colour are like ours Descartes claims that the blind man's ideas are different from ours. A seeing person assumes that her idea of red is like that of other seeing people because when she points at red objects, she can confirm that other people also call these objects red. By contrast even if a blind person would have an idea of red, he would not know it is the same as that of a seeing person. The blind

person is not able to confirm that her idea is the same as that of other people because she cannot see what others point at when they use the term 'red'. Thus, there is an aspect of the idea 'red' missing for the blind person. This aspect can be supplied only by the senses.

Other texts suggest that Descartes places even more importance on the sensory input. In the *Rules for the Direction of the Mind, Rule 14* he seems to require that we see at least the primary colours, and in the *Letter to Hyperaspistes* he continues: " ...even though the mind is indivisible, it is none the less *capable of acquiring* various properties" (CSMK III, p. 196). In my interpretation Descartes says here that the mind always had the capacity to acquire knowledge about red, but only when it is exposed to the sense experience of red does it acquire the new property of knowing red. This reading would support two conclusions. First, the content of the idea 'red' is not entirely innate but at least partly acquired through the senses. Second, the innate capacity that allows us to acquire the knowledge of 'red' from sense experience is not different from the innate capacity that allows us to acquire "pain, smell, and taste, and so on" (CSM I, p. 218).

Presumably, Descartes would not object to including 'sound' in 'and so on'. And, given what he says about our acquisition of words elsewhere ("When for example on *hearing* that the word "K-I-N-G" signifies supreme power, I commit this to my memory and then subsequently recall the meaning by means of my memory" CSMK III, p. 336, emphasis added), it appears plausible to assume that the same general-purpose faculty of sense perception contributes to the acquisition of the content of many of our ideas. In addition Descartes mentions here the faculty of

'intellectual memory', which is again a general-purpose faculty.

Again, my reading of the text does not find any indication that Descartes might have endorsed a view that language acquisition relies on a domain-specific subsystem of the mind. I have not been able locate any text that would support Chomsky's claim that Descartes believed "human intelligence [is] a specific biological system" (Chomsky, 1980, p. 7). On Descartes' view intelligence is not a biological but a mental system. And the Cartesian mind is, as Chomsky correctly acknowledges, "not a part of the biological world… [and] has not in itself any diversity of parts" (Chomsky, 1980, p. 30). Descartes' doctrine of innate ideas does not contradict his view about the indivisibility of the mind and does not give any indications that Descartes believes intelligence could be located outside of the mind. I have not found any indications that Descartes' view entails the 'irrational dogmatism' Chomsky attributes to him. Therefore, I suggest that there is no reason to ascribe to Descartes a view that he clearly did not hold.

While I have shown some important differences between Descartes' and Chomsky's view, one final task is left. Given that there are important differences between the acquisition of sensory ideas (such as red) and the acquisition of semantic concepts, I still need to show how the Cartesian mind acquires semantic concepts. Below I will inquire whether or not Descartes held that the acquisition of these concepts requires principles and/or faculties that are different from those required for sensory concepts. In order to do that I turn to two arguments that are frequently employed to support linguistic nativism: poverty of the stimulus arguments and impossibility arguments. The important question is whether or not Descartes was

committed to these kinds of argument regarding language acquisition.

2.4.2. Poverty of the Stimulus Arguments

Cowie (1999) holds that many defenders of innatism support their view with poverty of the stimulus arguments and/or with impossibility arguments. Both arguments presume that stimulus driven language learning would require that all knowledge of language must be in some form available in the stimulus received by the language learner (the primary language data, PLD). Both arguments also rely on the empirical claim that there is an insurmountable gap between the stimulus and the acquired language. Because the available input and general-purpose stimulus-response learning mechanisms cannot explain the output, proponents of both kinds of arguments conclude that language learning relies on some innate structures and/or knowledge. The important difference is that (i) poverty of the stimulus arguments imply that the required stimulus is, at least in principle, available in the PLD but that it is in fact not accessible while (ii) impossibility arguments imply that the required stimulus is not available in the PLD and could not have been learned even by a hypothetical child who had access to the complete corpus of a given language. In other words, the poverty of stimulus claim is that even though certain constructions are present in the complete corpus of PLD, they are so rare that not every child would encounter them. But virtually every child acquires the ability to produce these constructions. Therefore, language acquisition must depend on some form of innate knowledge and/or innate mechanisms. This means that poverty of the stimulus

arguments allow for the possibility that at least some linguistic knowledge is acquired from the input (available stimulus). If empirical research reveals that the input is rich enough, then potentially all linguistic knowledge could be acquired from this input. In Cowie's words:

> Poverty of the stimulus arguments grant, that the empiricist can succeed in explaining certain central cases of concept and belief acquisition. Their point is just that since some of the items we acquire are beyond the reach of the empiricist's general-purpose learning mechanisms we must view the mind as inherently containing additional, task-specific faculties. (Cowie, 1999, p. 49)

By contrast impossibility arguments imply that the input is not only impoverished in fact but in principle. This means that it would be impossible to acquire linguistic knowledge even if a child had access to the complete set of PLD. For arguments of this type it is irrelevant what we might find out about the richness of the actual input because no amount of input could explain the output. Essentially, the conclusion of impossibility arguments is that *all* linguistic knowledge must be innate: "impossibility arguments claim to establish that ... empiricism ... rests on a picture of mind-world interaction that is fundamentally incoherent or self-undermining ... everything that is in our minds is innate" (Ibid.)

Cowie proposes that Descartes defended both: poverty of the stimulus and impossibility arguments. Below I will examine her claims and show that Descartes was committed to neither argument regarding language acquisition. As a preliminary observation I submit that, while it may make sense to accept *either* poverty of the stimulus *or* impossibility arguments, it appears implausible that someone would

subscribe simultaneously to both. If all linguistic knowledge were innate, it would be irrelevant what type of/how much input a language learner receives.

According to Cowie (1999) Descartes provides explicit poverty of the stimulus arguments for the innateness of mathematical ideas. In *Meditation Three* he claims that it is impossible to form the idea of a chiliagon from sense experience and concludes that therefore this idea must be one of the 'true ideas which are innate in me" (CSM II, p.35). Throughout the Meditations this type of argument is applied to different kinds of mathematical ideas. Descartes seems to hold that even though we encounter simple geometrical forms (such as triangles) through our sense experience, these do not reveal the true essences of geometrical objects. And in the *Fifth Set of Replies* he explains:

> … no part of a line that was really straight could ever affect our senses, since when we examine through a magnifying glass those lines which appear most straight we find they are quite irregular and always form wavy curves…the true triangle is contained in the figure only in the way in which a statue of Mercury is contained in the rough block of wood. (CSM II, p. 262)

I suggest that this is not a true poverty of the stimulus argument. Descartes does not hold that 'true triangles' exist somewhere in the physical world but are so rare that they are inaccessible to most of us. He also does not seem to hold that the mind supplies us literally with a triangle with perfectly straight lines. Instead he seems to think that our faculty of rational thought allows us to grasp the true essence of triangles. This essence is what we could never obtain through observation of actual triangles, no matter how many triangles we might observe.

In the *Conversation with Burman* Descartes discusses the same problem from a different angle. Here he rejects the suggestion that we frame in our mind a perfect triangle from the imperfect triangles accessible to our senses. He argues that

> I could not conceive of an imperfect triangle unless there were in me the idea of a perfect one since the former is the negation of the latter. Thus, when I see a triangle, I have the conception of a perfect triangle, and it is by comparison with this that I subsequently realize that what I am seeing is imperfect... (CSMK III, p. 144)

However, Descartes does not tell us how the conception of the perfect triangle got into our mind. One possibility would be that it has always been there, that it is innate. Of course, this assumption pushes the question of where it came from just one step further: why does our mind contain a conception of something that does not exist anywhere in the 'outside world', and where is this conception coming from? For Descartes the final answer to these questions is: 'They have been placed there by God'. But this is not an answer that should be acceptable to Chomsky because it bars further inquiry. Nevertheless Chomsky recently gave an answer similar in flavour when he suggested that the creativity of language might be "one of those ultimate secrets that will ever remain in obscurity, impenetrable to human intelligence" (Chomsky, 2010a, p. 29).

If we do not want to accept that God is the ultimate cause of innate ideas we could propose that in our youth our imperfect senses fool us into seeing perfectly straight lines when in fact we see 'only' lines that are almost straight. Later when we discover our mistake, we have already formed (based on sense experience) the concept of a perfectly straight line. On this account, an innate idea of the 'true

triangle' would not be necessary for explaining why we have this conception. I do not want to claim that Descartes would be committed to the second account. For a discussion of Descartes' epistemology it could be important to rule out that this was his view. But it is irrelevant for my task. I am here only interested in Descartes' claims regarding language acquisition. And I do not believe that the two passages above provide any evidence for the claim that, as young children, we could not learn *the word* 'triangle' based on sense experience.

Seemingly, Descartes believes that knowing the word 'triangle' does not entail knowing the essence of triangles. In *Rules for the Direction of the Mind* he writes: "...I can have knowledge of a triangle *even though it has never occurred to me that this knowledge involves also knowledge of the angle, the line, the number three, shape, extension, etc*." (CSM I, p. 46, emphasis added), indicating that there is a difference between having *some* knowledge of a triangle (which presumably would suffice to use the word 'triangle' in appropriate circumstances) and understanding the essence of a triangle. In *Principles of Philosophy* Descartes seemingly allows an even more 'empirical' account of concept acquisition: "...when we *see* a figure made up of three lines we *form an idea* of it which we call the idea of a triangle; and we make later use of it as a universal idea..." (CSM I, p. 212, emphasis added). According to this passage, seeing an object (triangle) is sufficient not only for remembering that object later but also for forming a universal idea which we can apply later to objects of the same kind (other triangles). Similarly, when referring to numbers, Descartes claims that when considered in the abstract, "number ... is nothing more than a mode of thinking; and the same applies to all the other things which we call 'universals'" (I,

58; AT VIIIA, 27). Further, "universals arise solely from the fact that we use one and the same idea for thinking of all individual things which are similar to each other..." (I, 59; AT VIIIA, 27). This is an indication that our faculty of rational thought can abstract regularities from the input we receive. And for this faculty to engage it often seems sufficient to see only very few examples of triangles or numbers or other objects. However, such an account of acquiring general or universal concepts is quite compatible with current empiricist accounts of language acquisition.

As we have seen, Descartes also allows for different levels of knowledge of geometrical objects. And while the complete understanding of the essence of a triangle probably could not be learned from looking at (infinitely many) triangles, it appears not necessary to have such deep knowledge as a precondition for acquiring the word 'triangle'. It seems that Descartes allows that even more abstract concepts could be acquired before we have a complete understanding of their nature. For example in the *Fourth Set of Replies* he asserts that "...we can easily understand the genus 'figure' without thinking of a circle (though our understanding will not be distinct unless it is referred to some specific figure and it will not involve a complete thing unless it also comprises the nature of body)" (CSM II, p. 157). This passage seems to indicate that we need at least the experience with a small number of (different) figures before we can form the genus 'figure'. Furthermore, the contribution of any innate content seems to be limited to the most abstract level: we need reference to some specific figure to obtain distinct understanding, and we need understanding of the nature of body to have complete understanding. Again, for first language acquisition it seems the contribution of the innate content is not necessary.

I want to emphasize that I do not claim that Descartes would allow for a completely empirical account of knowledge acquisition or that he holds that we can come to understand the essence of things by purely empirical means. The evidence to the contrary is overwhelming. However, I am looking here only at first language acquisition. Specifically at the period of acquisition that occurs, according to Chomsky, faster and with greater ease than the acquisition of any other (intellectual) domain. For Chomsky these early accomplishments suggest the necessity of an innate domain-specific language faculty.

By contrast for Descartes innate ideas, as operations of the mind, are invoked to explain our capacity to understand the essence of substances. For this reason the poverty of the stimulus arguments in Chomsky are genuine in so far as it would be possible to argue that a child could learn language completely from empirical evidence if she just had access to the complete evidence. But, as a matter of empirical fact, she does not. (In chapter 4 I will provide empirical evidence challenging this claim). The alleged poverty of the stimulus arguments of Descartes, however, are not genuine. As I have shown above, Descartes does not claim that we could encounter empirical evidence for perfectly straight lines if we would carefully analyze all triangles, etc. Thus, his arguments are in fact impossibility arguments. In the next section I will inquire whether or not these impossibility arguments would commit Descartes to an account of language acquisition that foreshadows Chomsky's account.

2.4.3. Impossibility Arguments

When evaluating impossibility arguments we need to remember that Descartes uses the term 'innate idea' either to refer to the content of an idea (e.g., the idea of God "is an image of a true and immutable nature" (CSM II, p. 47)) or to operations of the human mind (e.g., by innate "we simply mean that we have within ourselves the faculty of summoning up the idea" (CSM II, p. 132)). I shall inquire here whether or not Descartes' beliefs about innate ideas would commit him to the belief in an innate language acquisition faculty as suggested by Chomsky. I will look at both aspects in turn: ideas individuated by content and understood as mental faculty.

Regarding content, Descartes discusses several cases in which this content cannot possibly come from outside of the mind; it cannot be transmitted from external sources. In these cases the missing content cannot even in principle be obtained through empirical sources (e.g., sense-perception). Descartes argues explicitly that some ideas (e.g., the idea of God) could not derive from the operations of our mind or from sense perception, and he concludes that they must have a different source. In the case of the idea of God this source is obviously God himself: "God, in creating me placed this idea in me…" (CSM II, p.35). Thus we all have the innate idea of God. But, as we will see below, we have very different degrees of access to this idea. I will now provide a detailed account of Descartes' discussion of the idea of God. I choose this idea for three reasons.

First, I consider it as a paradigm of an idea the content of which could not have been acquired from sense perception. In some of the other cases Descartes discusses (e.g., idea of a triangle) we may debate whether or not it would have been

impossible to derive the content of this idea from sense perception. By contrast it seems to be uncontroversial that Descartes believed we could not have acquired the idea of God from any source but God himself. For my inquiry it is not relevant whether or not Descartes has provided irrefutable proof for this belief, only that we have no reason to assume that he did not hold this belief.

Second, to support the view that the content of our ideas is innate, showing that at least the content of one idea is innate is sufficient. This is so because Descartes holds that all our ideas are connected: "...the items of knowledge that lie within the reach of the human mind are all linked together by a bond so marvelous, and can be derived from each other by inference.... (CSM II, p. 400).

Third, Descartes considers the knowledge of God as the foundation on which all other knowledge rests: "we must begin with the knowledge of God and all our knowledge of other things must then be subordinated to this single *initial* piece of knowledge" (CSM II, p. 290, emphasis added). Thus, if we could show that having the innate idea of God allows us to acquire language, we would have made some progress towards showing that Descartes was committed to an innate language acquisition device.

References to the idea of God are numerous in Descartes' writings. Many of these occur in connection with Descartes' attempts to prove the necessary existence of God (e.g., CSM I, p. 128; 197; 306; CSM II, p. 10, 28, 31, 35, 46, 76, 99, 162, 252,; CSMK III, p. 186, 194, 232, 331). While this topic is very important for Descartes, it does not concern me here. Rather, I am inquiring when and how we acquire the idea of God and whether or not it would be possible to derive this idea

from 'empirical sources'. Descartes explicitly denies the latter. In *Second Set of Replies* he writes:

> You suggest that I may have derived the idea, which gives me my representation of God from preconceived notions of the mind, from books, conversations with friends, etc. and not from my mind alone. But there is no force to this suggestion. If I ask these other people... whether they derive it from themselves or from someone else, the argument proceeds in the same way it does if I ask the same question of myself: the conclusion will always be that the original source of the idea is God. (CSM II, p. 98)

It is clear that Descartes denies that we could have gotten the content of the idea of God from sources other than God. However, the suggestions he considers and eventually rules out ("from books, from conversations with friends") concern sources of language acquisition. And Descartes does not deny anywhere that we could have learned (some of) the meaning of the word 'god' from these sources. This is especially evident when we analyze carefully an explicit claim Descartes makes regarding innate ideas in the infant's mind. In a *Letter to Hyperaspistes* Descartes writes "a mind newly united to an infant's body... has in itself the ideas of God, of itself, and of all such truths as are called self-evident, in the same way as adult human beings have these ideas when they are not attending to them; for it does not acquire these ideas later on as it grows older" (CSMK III, p. 190). This seems to be a claim very similar to some of Chomsky's claims regarding innate knowledge of language; the idea of God is already in the infant's mind, it is not acquired later.

However, on closer inspection it becomes evident that for Descartes the clear and distinct idea of God cannot always be identical with a given occurent idea of

'god'. We encounter again the ambiguity of the term 'idea'. It is undoubtedly true that Descartes holds that the true, clear and distinct idea of God is innate, put 'into' our minds by God. The most explicit statement to this effect occurs in *Meditation Three*: " God, in creating me ... placed this idea in me to be, as it were, the mark of a craftsman stamped on his work" (CSM II, p. 35). This passage could refer to both: content and mental capacity. I have already argued that for Descartes our species-specific mental capacity allows for language acquisition. And it is the content of the idea of God that we need to attend to when we seek to understand (as far as that is humanly possible) God. However, Descartes also allows that we can have a different idea of god. In *Comments on a Certain Broadsheet* he refers to some people (atheists) who believe that "all we can understand about God is what he is called, namely 'God', or what corporeal forms painters use to represent him" (CSM I, p. 305). Again, Descartes refers to the content of the "god" idea. And while he believes that the atheists are utterly mistaken about the essence of God he does not imply that they merely misunderstand the word 'God'. In fact, it is essential that the atheist understands (at least to some degree), what he denies. Thus, the occurrent idea of the atheist stands in some relationship to the clear and distinct idea of God and is not an entirely different idea. Descartes offers some explanations for why we could be mistaken about the content of innate ideas (e.g.,) but an adequate discussion of these would lead to far afield. The important point for my inquiry is that on Descartes account having the innate idea of God placed in our minds by God does not guarantee that we have the kind of 'automatic' access to this idea that Chomsky suggests children have to innate syntax and semantics. Mentioning the word 'God" or even

reading the Meditations will not 'trigger' the clear and distinct idea of God.

In *Discourse on Method* Descartes observes that "...many are convinced that there is some difficulty in knowing God... The reason for this is that they never raise their minds above things which can be perceived by the senses" (CSM I, p. 129). Here it appears as if Descartes claims that sense experience is not only insufficient for understanding God but also utterly irrelevant to such an understanding. While this might commit Descartes to some extreme innatism regarding the content of the idea of God, it appears to be a very different type of innatism from Chomskyan innatism regarding language acquisition. Furthermore, there is no indication in the passage that Descartes would deny that 'people who never raise their mind above things that can be perceived by the senses' are able to understand the word 'God' and use it in appropriate ways. In *Principles of Philosophy* Descartes admits that, unless we focus our attention on the contemplation of God, we can be misled by the "habit of distinguishing essence from existence...[and doubt] whether the idea of God is not one of those which we made up at will, or at least one of those which do not include existence in their essence" (CSM I, p. 198). This indicates that having the innate idea of God does not guarantee having a correct understanding of the concept. Yet, again, Descartes expresses no doubt regarding the ability of anyone to use the word 'God' appropriately. Clearly, the way in which the atheist or the inattentive believer uses the word 'God' is different (in kind) from the way in which the magpie uses the words 'good day'.

We have seen that Descartes allows that at least some people can understand the word 'God' without having (access to) a clear and distinct idea of God. This is the

case even though their minds contain the innate idea of God, as the following passage

from a *Letter to Cleselier* explicitly shows:

> ...although the idea of God is *imprinted on the human mind in such a way*
> that everyone has within himself the power to know him, that does not
> prevent many people from passing through their whole lives without ever
> having a distinct representation of this idea. Indeed, those who think they
> have an idea of many gods have no idea of God at all. (CSMK III, p. 248,
> original emphasis)

Apparently this passage contains a contradiction; on the one hand 'the idea of

God is imprinted on every human mind' yet on the other hand many people 'have no

idea of God at all'. To avoid this contradiction we have to recognize that Descartes

shifts his usage of 'idea' from 'power to know' to 'distinct representation'. But this is

not enough. We also need to acknowledge that the 'power to know' does not

necessarily entail that people have 'distinct representations'. In other words, on

Descartes' account it is possible to have a mistaken representation of God in spite of

having an innate power to know God. Descartes explicitly describes this possibility in

the same letter:

> When the ancients referred to many gods they did not mean many all
> powerful gods but only many very powerful gods, above whom they
> imagined a single Jupiter as sovereign; and consequently to this Jupiter
> alone they applied the idea of the true God, this idea being presented to
> them in a confused matter. (Ibid.)

Here Descartes offers a possible explanation for why the ancients, is spite of

having the innate idea of God, could hold a polytheistic view. First, on careful

inspection it becomes clear that not all ancient Gods were equally powerful. There

97

was one sovereign God who had more power than all the others and only this God was the 'true' God. Second, the idea of the true god was not presented clearly and distinctly to the ancients but 'in a confused matter'.

All of the foregoing is very different from Chomsky's various accounts of language acquisition. For Chomsky the innate language acquisition faculty is what allows children to acquire language effortlessly. For Descartes the possession of innate ideas does not guarantee the acquisition of clear and distinct concepts. Chomsky suggests that a child will know upon first encountering an expression containing a never heard before grammatical construction how to interpret this construction (e.g., Chomsky, 1975, 1986, 2002). Descartes allows that an atheist continues to remain ignorant about the true nature of God even after hearing 'God' many times, after debating the idea 'God' with others etc. For this reason I surmise that the content of Descartes' innate ideas plays an insignificant, if any, role for the acquisition of language.

The following analogy may illustrate how different the situation for Cartesian and Chomskyan language learner is. We can imagine two infants to be in two completely dark rooms. These imaginary rooms contain the content of the idea of God or all potentially possible grammatical constructions of a certain type (e.g., forming of polar interrogatives). Both infants receive some linguistic input. We can predict that at one point in the Chomskyan Room a light-switch is turned and the entire room is lit. No matter which of the innumerable constructions in the room the child encounters, she will unerringly know how to interpret it. The situation in the Cartesian Room is very different. A small candle is lit that allows the infant to sense

something in the dark. This is comparable to her "acquiring" the word 'God'. We cannot predict what will happen next. There are many other candles in the room and if the infant continues to light candles, one day she may grasp the full concept of God in all its divine glory. But there are no guarantees that this will ever happen. There is another important difference between the two rooms. Every infant who enters the Chomskyan Room will have the same experience: at one point the light turns on, and the room is entirely lit. By contrast some infants who enter the Cartesian Room will continue to light new candles and progress towards a clear and distinct idea of God. Others will light only very few candles and never progress beyond a confused or obscure idea of God. Thus, while for Chomsky the acquisition of grammar is effortless and uniform across the species, for Descartes the full acquisition of semantic concepts requires effort and varies considerably among individuals.

Turning now to faculties of the mind, we find textual evidence that for Descartes our ability to reason is innate. For example in *Third set of Objections and Replies* he states explicitly, "by innate we simply mean that we have within ourselves the faculty to summon up the idea…" (CSM II, p. 132). In a *Letter to Ceselier* Descartes asserts that innate ideas are "imprinted on the human mind in such a way that everyone has within himself the power to know..." (CSMK III, p. 248). In a *Letter to Hyperaspistes* he writes that "everyone has within himself an implicit idea of God ... an aptitude to perceive it explicitly" (CSMK III, p. 194). In *Comments on a Certain Broadsheet* he addresses the point several times, stating that "everything...is represented to us by means of ideas which come to us from no other source than our own faculty of thinking" (CSM I, p.304) and that "...by innate idea I have never

99

meant anything other than...that there is present in us a natural power which enables us to know..." (CSM I, p. 309).

An uncharitable reading could interpret these statements as contradicting what Descartes says elsewhere about the content of innate ideas. A more charitable reading will acknowledge that Descartes offers at least some suggestions regarding how one can reconcile the apparently inconsistent claims that (i) 'innate idea' refers only to a 'natural power' or 'faculty of thinking' and (ii) that innate ideas are separated from adventitious and fictitious ideas by the fact that their specific content (e.g., 'God' or mathematical truth) could not have been derived from sense experience or operations of the mind performed on other ideas (Williams, 2005). For example in a *Letter to Voetius* Descartes asserts that "all those things whose knowledge is said to be implanted in us *are not for that reason expressly known by us*; they are merely such that we come to know them by the power of our own native intelligence, without any sensory experience" (CSMK III, p. 222, emphasis added). Here Descartes suggests that the fact that (the content of) an idea is innate in and of itself does not guarantee that we have explicit knowledge of it. The same point is repeated in a *Letter to Hyperaspistes*

> … everyone has within himself an implicit idea of God, that is to say, an aptitude to perceive it explicitly; but I am not surprised that not everyone is aware that he has it or notices that he has it. Some people will perhaps not notice it even after reading my *Meditations* a thousand times. (CSMK III, p. 194)

We see again how different this notion of 'innate' is from Chomsky's: for Chomsky the innate language faculty allows children to learn language effortlessly

often even without being exposed to examples of what they learn. For Descartes it is possible to have an innate idea of God without becoming knowledgeable of it even after reading the '*Meditations'* a thousand times. If that is Descartes' belief (and I have no reason to doubt that it is), then having an innate idea God can play no role in language acquisition.

As I have previously shown, for Descartes the faculty of thinking is a unified faculty that underlies all mental activity. Thought is the essence of mind just as extension is the essence of body. Even in the application of hyperbolic doubt during *Meditation One,* Descartes never doubts his ability to think rationally. It would be impossible to doubt systematically if the 'operations of the mind' could not been taken for granted. So while Descartes conceives of the dream argument to doubt sense perception and of the deceiving God argument to cast doubt on mathematical truth and maybe even God, he does at no point doubt the ability to think rationally or to use language. If Descartes held that language is underwritten by a domain-specific faculty, it would appear odd to exclude language from the application of hyperbolic doubt without giving any justification for such an exclusion. If Descartes held (as I have suggested throughout) that language is a domain-general outward expression of thought, then no justification for its exclusion from doubt is needed.

There is one additional reason to suggest that Descartes did not believe that language depended literally on innate ideas. Ultimately, innate ideas must have their origin in God. This is true for both: operations of mind and objects of mind. Both are jointly necessary and sufficient for our ability to obtain knowledge. Descartes firmly held that God is the foundation of all knowledge. He wrote that "... knowledge is

conviction based on a reason so strong that it can never be shaken by any stronger reason. Nobody can have the latter unless he also has knowledge of God" (CSMK III, p. 147). It is here not relevant whether or not Descartes proved beyond doubt that God is the ultimate foundation for knowledge. It is important, however, that Descartes *believed* that God was this foundation.

Now, Descartes never went so far as to suggest that innate ideas guarantee true judgment. He believed in free will, and he allowed for human error. But Cartesian optimism, the conviction that we *can* obtain true judgment, depends on the belief that we have the mental tools that are sufficient for knowledge acquisition. Descartes believed that our minds are able to know (because they are created by God who is not a deceiver) and that our minds are furnished with some content that we could never obtain through sensory experience (ideas of God, mathematical truths etc.). Several commentators (e.g., Schmaltz, 1997; Flage & Bonnen, 1999) have shown that Descartes was careful to avoid the implication that God is the author of all of our thoughts. But while Descartes allowed for the possibility that some people would "[pass] through their whole live without ever having a distinct representation of [the idea of God]" (CSMK III, p. 248), he also held that we all have within ourselves "the power to know him" (Ibid.). And once we become distinctly and clearly aware of the content of innate ideas (e.g., the Pythagorean theorem of right angled triangles), we cannot 'go back' and deny that "... a triangle having the square on its hypothenuse equal to the squares on the other sides ... is right-angled" (CSMK III, p. 158).

Finally, looking specifically at language, Descartes indicates nowhere that the

grammar and vocabulary of human language is similar to mathematical truths or the idea of God. In his only lengthy discussion of language (*Letter to Mersenne*) he observes that our languages contain "defective or irregular verbs introduced by corrupted usage" (CSMK III, p. 11), "that all the differences in inflexion of words have been introduced by usage ... [because] what is pleasant in our language is coarse and intolerable to Germans and so on" (CSMK III, p. 12) and that "almost all our words have confused meanings" (CSMK III, p. 13). While Descartes seems to accept that it is, at least in theory, possible to invent a universal human language, he does not suggest it would be based on uncovering and/or consciously accessing the innate principles of language. Instead he proposes that this hypothetical language should be based on principles similar to those of mathematics: "Order is what is needed: all the thoughts which can come into the human mind *must be arranged* in an order like the natural order of numbers" (CSMK III, p. 112, emphasis added). Here it becomes clear that Descartes does not believe that our languages are already arranged in a 'natural system' similar to that of numbers.

Next, Descartes suggests that the 'discovery' of a universal language depended on 'the true philosophy' which allows us "to number and order all the thought of man ... to separate ... [and] explain correctly what are the simple ideas in the human imagination out of which all human thoughts are compounded" (CSMK III, p.13). This passage suggests that Descartes believes that if we could ever hope for a universal language it would depend on knowledge that we acquire through 'true philosophy'. Since this language would have to be actively 'constructed', it could not be innate. While it is possible to maintain that the knowledge on which the universal

language is based would be innate, it appears that the vocabulary and grammar of this language still would have to be *taught* to children and would not be innately available to them when they first acquire language. And Descartes remains very skeptical about the prospects of such a universal language: "I do not hope to ever see such a language in use. For that *the order of nature would have to change* so that the world turned into a terrestrial paradise; and that is too much to suggest outside of fairyland" (Ibid., emphasis added). While Descartes does not elaborate what he means here with 'the order of nature would have to change', it seems clear that *if* our current language would depend on innate knowledge, then this innate knowledge (as part of the current order of nature) would have to change to allow for a language that "represents matters so clearly that it would be almost impossible to go wrong" (Ibid.). And this would imply either (i) that God furnished our minds with some innate knowledge (language) that is not, even in principle, suited for the acquisition of true knowledge (he was a deceiver after all) or (ii) that Descartes did not believe that the grammar and vocabulary of language depend on innate knowledge. Since defending (i) entails the claim that Descartes was insincere about much of what he said about the dependence of our knowledge acquisition on a non-deceiving God and since Descartes at least on one occasion explicitly states that languages can be acquired based "on experience alone" (CSM II, p. 403), I suggest that we have little reason to believe that he held language relied on innate ideas.

2.5. Conclusions

In this chapter I have provided textual evidence supporting my belief that Descartes would not have agreed with several core claims of Chomsky's *Cartesian Linguistics*. Specifically, regarding Chomsky's claims that language is species specific, domain-specific and depends on an innate biological language faculty, Descartes would only be committed to the first. He held that language is "a real specific difference between humans and animals" (CSMK III, p. 366). One point he stresses repeatedly is that even the dumbest child outperforms the most talented animal when it comes to language. Like many contemporary philosophers, anthropologists, psychologists, linguists, and primatologists, he believed that the difference between human language and animal communication systems is one in kind not in degree. This very fact makes it implausible that he would hold that the acquisition of language is mechanical (as proposed by Chomsky). He believed that some words could be acquired 'mechanically' (e.g., by animals) but denied that the resulting abilities we are the same level as human language:

> If you teach a magpie to say good-day to its mistress when it sees her approach, this can only be by making the utterance of this word the expression of one of its passions. For instance it will be the expression of the hope of eating if it has always been given a tidbit when it says it. (CSMK III, p. 303)

This purely mechanical process of vocabulary teaching is achieved by 'exploiting' the passions of an animal that does not have a mind. The magpie does not learn words of our language but imitates a whole phrase because it is rewarded for this imitation. And if we stop feeding the magpie tidbits, it will eventually stop saying

'good-day' when it wants a tidbit. On Descartes' view the magpie does not have a little bit of language when she says 'good-day', she has *no* language at all. For Descartes the purely mechanical process of 'teaching' the magpie words is comparable to teaching a dog a trick that "can be performed without any thought" (CSMK III, p. 303). It is not at the same level as teaching an infant her first words because language learning requires a mind.

I have found virtually no evidence supporting the claims that Descartes thought language is domain-specific and that it depends on a specialized language faculty, which is largely independent of 'general intelligence'. Descartes holds that even people with exceptionally low intelligence (idiots, madmen, stupid children) have language. But he never indicates that the quality of their language use is not affected by the low level of their intelligence. But he has very little to say about the differences between individual human speakers (other than acknowledging their existence, e.g., CSMK III, p. 366). Descartes suggests that language is a reliable indicator of thought and that thought in turn depends on the rational mind. For Descartes minds are indivisible, but they can acquire different properties (CSMK III, p. 196). This allows for learning from experience. Language is one but not the only reliable indicator of thought. Descartes also holds that humans (but not animals) are capable of responding in appropriate ways to innumerable different situations (action test, CSM I, p. 140). Finally, Descartes does not limit thought to linguistic thought but considers willing, imagining, and even sense perception as forms of thought. It is possible, but not necessary, that language is involved in all forms of thinking. Furthermore, since Descartes holds that the mind never is without thought and that

pre-linguistic infants have minds, it follows that Descartes does not hold that all thought depends on language. Rather he believes that the general-purpose faculty of thought underwrites language.

I have not found confirmation for the claim that Descartes believed language acquisition depends on some form of innate knowledge that is comparable to the knowledge provided by the Chomskyan language faculty. Descartes uses the term 'idea' ambiguously, referring to either objects or operations of mind. Considering ideas as objects of mind, it seems that if Descartes' innate ideas would affect language acquisition, they (only) would affect the acquisition of semantic content. Regarding grammar Descartes mentions only in passing that it is learned (CSMK, III, p. 10) but he never explains how this learning occurs. And when briefly discussing the many problems with existing languages, Descartes never suggests that we could uncover a 'true' grammar within us in a similar way as we could discover the true idea of God or the true essence of a triangle. Instead, he suggests that humans would need to reform existing grammar or invent a new one that could be applicable to all languages (CSMK III, p. 11). This suggests that the rules of grammar cannot be innate.

The role that innate ideas play in the acquisition of semantic content for Descartes is very different from the role that the innate language acquisition device plays for Chomsky. On the one hand Cartesian innate ideas are 'born with us' (CSM I, p. 304) and 'not acquired later' (CSM III, p. 190). On the other hand having an innate idea of God does not entail that an individual has explicit and complete knowledge of God. Even after hearing the word 'God' many times, after reading the

Meditations thousands of times (CSM III, p. 194) it remains possible that an atheist has no clear and distinct grasp of the true idea of God that is innately in his mind. Similar points hold for other innate ideas (mathematical truths, self, etc). This indicates that for Descartes the semantic content of innate ideas cannot play any significant role in first language acquisition.

When considering ideas as operations of mind (or capacities, potentialities), we find that on the one hand it is obviously true that 'innate ideas' allow for language acquisition. But this is rather uninformative because it does not say more than: humans can learn language because they have some innate capacity to learn language. We already knew that humans (but not animals) can learn language. We also know that there are other differences between humans and animals (culture, technology, art, etc.), but we do not seem to hold that these depend on domain-specific innate operations of the mind. On the other hand it is very problematic to suggest that Descartes holds that some operations of the mind are *exclusively* dedicated to language acquisition. Given that he holds that "the ideas of pain, colors, sound, and the like must ... be innate" (CSM I, p. 304) it is difficult to maintain, in the absence of any explicit argument by Descartes, that language depends on innate ideas (capacities) that are exclusively specific to the faculty of language.

Finally, the fact that Descartes holds that innate ideas ultimately originate from God is difficult to reconcile with the fact that he also acknowledges the (logical) shortcomings of the grammar and vocabulary of existing languages ("defective and irregular verbs", "confused meanings"). The Cartesian God is no deceiver; for Descartes all error is human. And while Descartes suggests in many other cases that

error can be overcome by paying close attention to the 'clear and distinct ideas' we find in us, he never suggests such a 'fix' for language. He seems to hold that we learn our language from "books for practice in reading [and] speakers for practice in conversation" (CSMK III, p. 12).

None of the arguments I have provided above supplies by itself unequivocal support for my claim that Descartes did not believe language acquisition relies on innate ideas. But taken together they challenge Chomsky's suggestion that his views about language acquisition and language use can be traced back to a Cartesian tradition, if the term 'Cartesian' has anything like its commonly accepted meaning. The evidence I evaluated suggests the possibility that Descartes would endorse a view of language that combines components of current empiricist approaches to language acquisition (as discussed in chapters 4 and 5) and 'rational realism' proposed by Katz (1998). More research is needed to confirm whether such a combination view would be viable and whether it would be Cartesian in spirit. Chomsky (1966) was correct to suggest that a rich linguistic tradition exists. A rediscovery of these ideas may indeed provide valuable impulses for contemporary research. But in order to re-discover the capital of Cartesian ideas we need to pay close attention to the writings of historical figures and resist the temptation to infuse our own thoughts into their views.

Chapter 3. The Evolution of Chomskyan Linguistics

3.1. Introduction

According to many commentators Noam Chomsky's work has revolutionized linguistics and it continues to affect debates in the philosophy of language. The pervasiveness of Chomsky's ideas is demonstrated by their acceptance in a wide range of texts (e.g., Boeckx & Hornstein, 2010; Brook & Stainton, 2000; Cattell, 2006; D'Agostino, 1986; Fodor, 2000; Hornstein, 2005; Hymers, 2005; Kasher, 1991; Leiber, 1975; Lightfoot, 2005; Martin, 1987; McGilvray, 1999, 2002, 2005; Pietroski & Crain, 2005; Russell, 2004; Smith, 1999; Stainton, 1996). Chomsky's promise to bring rigour and exactness to linguistics (Chomsky, 1951, 1957) and to situate linguistic theorizing firmly within the biological sciences (Chomsky, 1965a, 1966) clearly revived linguistics and resulted in extensive research projects (e.g., Harris, 1993; Seuren, 1998; Boden 2006; Newmeyer, in press). His influence has inspired commentators to pen descriptions like "cognitive revolution" (Otero, 1988, p. 14), "Chomskyan revolution" (Lyons, 1970, p. 1; Harris, 1993, p. 35), "Chomskyan Turn" (Kasher, 1991, p. 4; Winston, 2002, p. 15) and to call Chomsky the "originator of the cognitive revolution" (Smith, 1999, p. 4). It has been acknowledged that "Chomsky's theory undeniably occupies center stage in … linguistics [and is] by far the best worked out" (Katz, 1981, p. 11). Chomsky's work has been called "a science in the Cartesian-Galilean tradition" (McGilvray, 2005, p. 4) that is "gaining [Chomsky] a position in the history of ideas on par with that of Darwin or Descartes" (Smith, 1999,

p. 1) and he has been acknowledged as the "Einstein of linguistics" (Leiber, 1975, p. 19). His writing has been called "majestic" (Mukherji, 2010, p. 2) and is considered seemingly of biblical proportion: "In the beginning there was *Syntactic Structures*" (Boeckx & Hornstein, 2010, p. 116). Commentators suggested "nothing has had a greater impact on contemporary philosophy than Chomsky's theory of language" (Harman, 1974, p. vii) and that "Chomsky reconceptualized learning" (Gallistel, 2010, p. 193).

On the other hand Chomsky's work has been massively critiqued by scholars of various fields (e.g., linguistics, philosophy, developmental psychology, cognitive science, language evolution) to the point where some have asserted that "…the overall quality of [Chomsky's] linguistic theorizing…and descriptions in terms of generative grammar has declined dramatically" (Seuren, 1998, p. 526), that "in some ways… [Chomsky] *set back* the field of linguistics and *undermined* its relevance for cognitive science" (Boden, 2006, p. 592, original emphasis) and that after 60 years of research "the Chomskyan hypothesis of an innate universal grammar… has no coherent formulation" (Tomasello, 2009, p. 312).

In this chapter will focus on Chomsky's work and use the term 'Chomskyan' as referring to him or to his close followers. It is important to keep this limitation in mind because over the years many linguists and cognitive scientists who were at one time closely aligned with Chomsky have developed views that deviate more or less substantially from his view. It would go well beyond the scope of this dissertation to follow all these developments. My limitation in scope is justified because "…for three decades and more the academic stage [of linguistics] has been dominated by the

111

imposing persona of Noam Chomsky" (Huck & Goldsmith, 1995, p. viii), the "Einstein of linguistics" (Leiber, 1975, p. 19). Chomsky's work, rightly or wrongly, is often used as *the* measuring stick for linguistics. Hence, I consider it important to be clear about what the commitments of Chomsky's view are. It should be also noted that by focusing on Chomsky I focus on metalinguistic issues because "[most] of Chomsky's work in theoretical linguistics deals with metatheoretical questions, which makes Chomsky a metalinguist rather than a linguist" (Seuren, 1998, p. 252).

I will show that, over the years, the label 'Chomskyan' has become problematic for several reasons. First, several linguists and philosophers who contributed greatly to Chomsky's research project in the early 1960s have drifted away from Chomsky and developed competing linguistic views. As a consequence some of their contributions have not been appropriately acknowledged by Chomsky (e.g., McCawley, Lakoff, Postal, Ross; For discussion see Harris, 1993; Seuren, 1998; Levine & Postal, 2004; Murray, 2010). Second, Chomsky holds a complex of closely related but logically independent commitments. Many linguists who are referred to as Chomskyan today share only some of his commitments but diverge, at times substantially, from others (e.g., Jackendoff, Partee, Stainton, Wilson). This makes it difficult to apply the term 'Chomskyan' consistently to a group of linguists. Third, Chomsky's own views have undergone many substantial revisions that are not always clearly acknowledged. Thus, even when applied strictly to Chomsky's own views, the label 'Chomskyan' is based on the fact that these views have been held by the same person and does not refer to a coherent body of linguistic theorizing. I will only focus on Chomsky's theorizing and trace it's development. McGilvray's outline in the

introduction of the third edition of *Cartesian Linguistics* will provide the broad framework.

I introduce Chomsky's view and show how it has evolved over time. Special attention will be paid to his arguments for the existence of a genetically determined domain-specific language acquisition device (LAD). In this context I will discuss the importance of clear definitions of key terms like 'innate', 'Universal Grammar' (UG), 'Language Faculty' (LF), and 'language acquisition device' (LAD) and show that Chomsky has not consistently used these concepts. This has resulted in unnecessary confusion and I will suggest that we are in need of clarifying work, especially for the theorizing from the 1990s to present. Given that Chomsky promised to place linguistics in the natural sciences, I will discuss his view of scientific inquiry and show how it has shifted over the years. I conclude that recent commentators are correct to assert that Chomskyan linguistics no longer meets widely accepted criteria for scientific theorizing (e.g., Seuren, 1998, 2004; Levine & Postal, 2004; Tomasello, 2005; Boden, 2006; Newmeyer, 2008). Further, I will situate Chomsky's arguments within the rationalist-empiricist debate and show that they do not support the conclusion that an innate, domain-specific language acquisition module is *necessary* to account for language acquisition. I will discuss the evolution of Chomsky's rationalist view and argue that it has not provided a satisfactory explanation for facts of linguistic creativity and for language acquisition. My findings stand in stark contrast to McGilvray's assertion that Chomsky's theorizing has led to "very considerable progress on all fronts" (McGilvray, 2009, p. 22). I shall begin with sketching McGilvray's view and then show in sections 3.3 – 3.5 why it is incorrect.

In the introduction to the third edition of *Cartesian Linguistics* McGilvray describes two contemporary research strategies to study the mind and language: 'empiricism' and 'rationalism'. He holds that these strategies are mutually exclusive and combinatorily exhaustive. According to McGilvray, Rationalist Romantics (RR) "hold that the mind's concepts and the ways of putting them together in language and thought are largely innately configured" (p.6). RR researchers are innatist, internalist, and nativist, and this combination allows them to account for 'everyday linguistic creativity', which is acquired by children at an early age (four years according to McGilvray, p. 7). Innate concepts alone can account for the uniform acquisition of language across human populations in spite of poverty of stimulus facts, or so McGilvray claims. On this account the child does not learn language but accesses what is innately available to her: "...the mind's concepts and the way of putting them together in language and thought are largely innate" (p.6), "the only way to explain the early appearance of creativity is to assume innateness of both concepts and combinatorial principles" (p. 7); "Innateness provides a basis for understanding one another even at a young age" (Ibid.); and "concepts and language are somehow implicit in some kind of natural 'mechanism' of the human body-mind, under (partial) control of the genome and the course of development it controls" (p. 18f). McGilvray asserts that Chomsky's RR theory is simple, objective, and descriptively and explanatory adequate. It accommodates the science of language to biology and makes steady progress.

This enthusiastic endorsement of Chomsky's work stands in stark contrast to the sobering evaluation expressed here:

> …almost all of Chomsky's linguistic views have been controversial and have, we would suggest, become ever more so over time… claims and promises made during the early years of his academic activity…have over time largely proved to be wrong or without real content and the promises unfulfilled (Levine & Postal, 2004, p. 203)

In the following sections I will argue that Chomsky's RR theory does not live up to McGilvray's praise and show why the criticism of Levine and Postal is justified.

3.2. Linguistic Nativism/Rationalism

3.2.1. Situating Chomsky's Nativism/Rationalism

McGilvray's black-and-white view of linguistic theorizing does in no way reflect the complexities of actual linguistic research. While it is possible, to a very rough approximation, to classify researchers into 'empiricists' and 'rationalists', such a classification does not reflect the subtlety of individual views. Thus, it is desirable to paint a more nuanced picture. We do not need to start painting from scratch as several commentators have provided detailed proposals. For example, Fiona Cowie (1999) introduces a framework for evaluating different forms of rationalism (nativism in her terminology). For Cowie Chomsky's nativism is a complex view committing him to several claims that are logically independent. First, Chomsky endorses a form of representationalism. He holds that contentful mental states (representations) play a crucial role in the production and explanation of linguistic behaviour. Our minds contain a finite stock of simple ideas and rules for combining them into more

115

complex thoughts. Second, he endorses biological bondedness, which places restrictions on the kinds of contents our thoughts can have. Thus, our biology constrains the possible hypotheses that language learners entertain. Third, Chomsky endorses domain specificity, which limits the hypotheses children can entertain during language learning even further by principles that are specific to the linguistic domain. Fourth, Chomsky defends the innateness of language. He holds that the constraints on language learning are to a large degree innate (presumably encoded by our genes). Fifth, and probably most importantly, Chomsky introduced Universal Grammar, which identifies the nature of the domain specific language faculty. "Universal Grammar is a theory about those features that natural languages have qua natural languages. It describes the class of 'linguistic universals', the properties that are common to all languages" (Cowie, 1999, p. 156).

Cowie suggests that a multitude of nativist positions is possible and that most of them are not compatible with Chomskyan nativism. What characterizes Chomsky's view is his commitment to all five core claims: "Representationalism, Biological Bondedness, Domain Specificity, Innateness, and Universal Grammar together constitute … Chomskyan Nativism" (Cowie, 1999, p. 157). This evaluation is consistent with an earlier system by Fred D'Agostino. According to him, Chomsky defends a form of linguistic rationalism that commits him to linguistic mentalism, linguistic subjectivism, linguistic intellectualism, and (biological) limitationalism. This means Chomsky holds that every speaker of a language 'knows' implicitly the grammar of this language (mentalism) and that this grammar describes the psychological basis of linguistic competence (subjectivism). Our linguistic behaviour

is rule-governed (intellectualism) and "involves interactions between language users and their environment which are best described in rational, computational terms" (D'Agostino, 1986, p. 114). Furthermore, according to Chomsky, our linguistic development is critically limited by biologically fixed, innate cognitive faculties (biological limitationalism). Like Cowie, D'Agostino holds that the debate surrounding Chomskyan nativism often fails to distinguish between logically independent components of Chomsky's view. Hence it is important to distinguish Chomskyan nativism/rationalism from other forms of nativism/rationalism. In this chapter I will focus mainly on the implications Chomsky's nativism/rationalism has for his theorizing about language acquisition.

Cowie and D'Agostino emphasize the different forms of nativism/rationalism. Sverker Johansson (2005) places Chomsky's theorizing in a much broader framework. He surveys a wealth of nativist and non-nativist language acquisition theories and suggests that they can be classified as follows:

- empiricist theories (connectionism and probabilistic/distributional approaches)
- cognitivist theories (functionalist approaches, language emergent from cognition)
- social-cognitive interaction theories (socio-perceptual language emergence, cultural language acquisition, ecological language acquisition, context- and usage based language acquisition)
- nativism (Chomskyan theories)
- neo-nativism (optimality theory) (Johansson, 2005, p. 185)

For Johansson Chomsky's commitments include an innately fixed universal grammar, domain-specific cognitive constraints, representationalism, and biological boundedness. However, he also observes that especially the more recent theories of Chomsky (e.g., the Minimalist Program, 1995) greatly reduce the reliance on genetically specified innate structures. Here I will not evaluate these classification systems regarding completeness and accuracy, but only use them as rough guides for situating the views of Chomsky and some of his critics. The main interest of my work is language acquisition, and so I will focus specifically on innateness, domain specificity, Universal Grammar and the language acquisition device.

3. 2. 2. The Problem of Innateness

The problem of defining innateness is not new. In chapter 2 we have already seen that Descartes did not always use the term 'innate idea' consistently. David Hume pointed out that from different definitions of 'innate' quite different implications follow:

> If innate be equivalent to natural then all the perceptions and ideas of the mind must be allowed to be innate... If by innate be meant, contemporary to our birth, the dispute seems to be frivolous... But admitting these terms, impressions and ideas, in the sense above explained, and understanding by innate ideas what is original or copied from no precedent perception, then we may assert that all our perceptions are innate and our ideas are not innate (Hume, 1958, p. 21).

Unfortunately, inconsistency in using the term 'innate' has persisted in to

present day usage. Mameli & Bateson, (2006) surveyed the literature in an attempt to systematize concepts of 'innateness'. Their work shows that we are still far away from one widely accepted definition. These authors describe "26 different candidates for...[the] concept of innateness [and conclude that] none is problem-free" (p. 176). The list below is taken from their survey:

A trait is innate if and only if:

(1) It is not acquired.
(2) It is present at birth.
(3) It reliably appears during a particular stage of the life cycle.
(4) It is genetically determined.
(5) It is genetically influenced.
(6) It is genetically encoded.
(7) Its development does not involve the extraction of information from the environment.
(8) It is not environmentally induced.
(9) It is not possible to produce an alternative trait by means of environmental manipulations.
(10) All environmental manipulations capable of producing an alternative trait are abnormal.
(11) All environmental manipulations capable of producing an alternative trait are statistically abnormal.
(12) All environmental manipulations capable of producing an alternative trait are evolutionarily abnormal.
(13) It is highly heritable.
(14) It is not learned.
(15) (i) It is psychologically primitive and (ii) it results from normal development.
(16) It is not produced by developmental mechanisms adapted to produce different traits in response to different environmental conditions.
(17) (i) It is not produced by a mechanism evolved to map different environmental conditions onto different phenotypes and (ii) it results from normal development.
(18) (i) It is not produced by a mechanism adapted to map different environmental conditions onto different phenotypes and (ii) it does not result from the impact on development of evolutionarily abnormal environmental factors.
(19) It is generatively entrenched in the design of an adaptive feature.
(20) It is insensitive to some range of environmental variation.
(21) It is developmentally environmentally canalized, i.e. there exists an evolved mechanism adapted to ensure that the development of the trait is

robust with respect to some environmental perturbations.
(22) It is post-developmentally environmentally canalized, i.e. there exists an evolved mechanism adapted to ensure that the continuance of the trait is robust with respect to some environmental perturbations.
(23) It is species-typical.
(24) It is a Darwinian adaptation.
(25) It is a standard Darwinian adaptation.
(26) It is prefunctional. (Mameli & Bateson, 2006, p. 177).

Given this plethora of definitions, it is not surprising that we still encounter (at times significant) confusion about what it means to say an idea or a cognitive mechanism is innate. To illustrate some of the problems arising from the possibility of multiple interpretations I want to look here only at two of the frequently used definitions: (2) A trait is innate if and only if it is present at birth, and (14) A trait is innate if and only if it is not learned. This choice is motivated by the fact that Chomsky's use of 'innate' seems to imply that he relies implicitly on some variation of these definitions. I will show that even though the definitions seem precise, it is by no means always clear how to apply them.

When we say a trait is 'present at birth', we usually mean it is encoded in some way in the genome and has not been shaped by experience. However, there are several problems with this interpretation. While many inborn traits (e.g. sex, ability to breathe, sucking reflex, eye colour, etc.) are present at birth, others (e.g., secondary gender characteristics, certain genetically transmitted diseases, etc.) develop only at a later time. These traits develop virtually independently of external input according to a preset genetic program. Learning and other 'environmental factors' have only a negligible impact on the manifestation of these traits. Thus, it would be more accurate to say that either traits or dispositions to develop them are present at birth. For

120

Chomsky innate seems to mean that a trait (e.g., language) is determined by genetic factors even though it may not be present (completely developed) at birth (e.g., Chomsky, 1975a, 1980, 1986a, 2005, 2009a, 2010b).

There are, however, some problems with such a broad-brush account of 'innate'. For example, it is well documented that two medical conditions that are 'present at birth' but only 'develop fully' over time (Down Syndrome (DS) and Fetal Alcohol Syndrome (FAS)) have quite different causal histories. DS could not have been prevented by different behaviour of the expectant mother because it is genetically determined. On the other hand, FAS is a consequence of the excessive alcohol consumption of the mother during pregnancy. Such behaviour may occur more frequently in women who are alcoholics, but there is no known link to genetic causes. Thus, we would consider DS as an innate trait that develops 'independently of external input according to a preset genetic program'. However, FAS depends on non-genetic causes even though it is already present at birth. We will see in chapter 4 that these distinctions play an important role in language acquisition. At least some of the abilities that are already present at birth have been traced back to pre-natal exposure to language. This means that even though these abilities are 'present at birth', they have been shaped by experience. Since we are only beginning to explore what the unborn child can and does acquire from exposure to spoken language, it seems sensible to abandon an account of innateness that equates innate and 'inborn'.

We encounter similar problems regarding the definition that a trait is innate if it has not been learned. First, to oppose traits that are learned with those that are unlearned creates a category that does not contain only innate traits but also many

traits that we do not want to call innate. For example, if a person loses a leg in a car accident and gets an artificial limb, or if another person gets a pacemaker, we do not say that the artificial limb or the pacemaker are 'learned' traits. But they certainly are not innate either. These examples may create the impression that it will be easy to rule out cases that are not relevant to the 'innate' vs. 'learned' distinction. However, frequently we find intricate interactions between environmental factors, innate endowment and learning, and it will be often difficult to disentangle these factors.

Second, we encounter many different accounts of 'learning' in the literature. Chomsky mentions and dismisses only some of them: "[empiricists hold that language] is taught by 'conditioning' (as would be maintained, for example, by Skinner or Quine) or by drill and explicit explanation (as was claimed by Wittgenstein)" (Chomsky, 1965a, p. 59). This brief sketch fails to acknowledge the subtleties of Quine's and Wittgenstein's views. And it does not exhaust 'empiricist' accounts of learning. Some accounts have emphasized 'imitation' (e.g., Guillaume, 1971; for a recent overview see Speidel & Nelson, 1989) or observation (Bandura, 1977), others highlight, 'association' to analogous forms, induction from examples or theory construction (e.g., Bruner, 1966). However, Chomskyans have claimed that no 'empiricist' account could explain language acquisition without the stipulation of at least some innate knowledge (Chomsky, 1980; Smith, 1999; McGilvray, 2005, 2009; Catell, 2005, Pietroski & Crain, 2005). They hold that only some aspects of our language behaviour are acquired by imitating the words we hear or through explicit instruction or through other 'empiricist' teaching methods.

However, the fact that particular accounts that emphasize one particular aspect

of learning fail to give a complete explanation of language acquisition does not imply that language is not learned. Many recent accounts of learning (e.g., Moerk, 1992; McClelland, et al. 1995; Tomasello, 1992, 2003; Deacon, 1997; Redington & Chater, 1997; MacWhinney, 2004; Sampson, 2002; Reali & Christiansen, 2005) integrate several aspects of earlier accounts because during the last decades it has become evident that learning can occur in many situations which are very difficult to 'quantify'. This is especially true for learning that occurs pre-natally, during infancy and in early childhood. Currently the available evidence does not rule out that general-purpose mechanisms could account for language acquisition. I will suggest in chapter 4 that we need to account for all forms of learning that are involved in language acquisition no matter when and how they occur.

I do not believe it is desirable to add further definitions of 'innate' to the already existing abundance. Nor does it seem feasible to follow the suggestion of Scholz & Pullum (2006), Bateson (1991), Griffith (2002) and others that the "the term 'innate' should be abandoned in theorizing about language acquisition because it impedes the study of language acquisition" (Scholz & Pullum, 2006, p. 66). This certainly would resolve the many fruitless debates that frequently concern not even the same content. However, it is simply not realistic to expect that those engaged in the debates will give up the term 'innate' without having a widely accepted replacement. Thus, when we evaluate innateness-claims, we need to pay close attention to how the term is defined and whether or not this definition is applied consistently. In directing specific criticisms to specific definitions we may be able to eliminate eventually the more problematic cases. In addition it would be helpful to

follow Scholz and Pullum's advice to shift the focus from the alleged differences between innate and learned and attempt "a detailed examination of all mechanisms that play a role in language acquisition" (Scholz &Pullum, 2002, p. 67).

Given that one of the aims of Chomsky's work has been to situate linguistic theorizing within the biological sciences it may be expected that his work has focused on definitional issues for problematic terms like 'innate' and that he has developed a testable hypothesis concerning the innate mechanism that underwrite language. Surprisingly, this is not the case. We find repeated assertions that Chomsky has never defended an innateness hypothesis: "I have never defended [the innateness hypothesis] and have no idea what it is supposed to be" (Chomsky, 2000b, p. 66), "[the innateness hypothesis] … has never been formulated or defended" (Chomsky, 2003, p. 316); "there is no such general [innateness] hypothesis" (Chomsky, 2007 a, p. 1099).

What can explain the fact that so many critics and even supporters of Chomsky are mistaken and attribute an innateness hypothesis to him? First, there is a technicality. Chomsky seems correct to insist that he has never formulated and defended anything labeled *the innateness hypothesis* (even though he did discuss in some detail *the "innateness hypothesis"* in Chomsky, 1975a, pp. 33-35). This insistence could come at a price. It would confirm that, over decades, Chomsky has avoided abiding by commonly accepted scientific practice: he has never formulated a testable and potentially falsifiable hypothesis regarding the language faculty. Aware of this problem, Chomsky has repeatedly claimed that his critics (e.g., Putnam, 1988; Gopnik, 2003; Boden, 2006) have erroneously alleged that he was holding a *general*

124

innateness hypothesis when in fact he has formulated "*specific* hypotheses about the innate resources of the mind…its language faulty" (Chomsky, 2000a, p. 66, emphasis added).

The following examples show that it has not always been obvious why a proposed hypothesis is specific and not general: "we construct a second-order hypothesis as to the innate principles that provide this [linguistic] competence on the basis of presented data" (Chomsky, 1967, p. 2) and we need to "formulate a hypothesis about innate structure that is rich enough to meet the condition of empirical adequacy" (Ibid., p3). As we shall see in section 3.5., most of Chomsky's hypotheses about the nature of the language faculty are not any more specific. And at times it is even more difficult to maintain that Chomsky is *not* referring to a general hypothesis. For example "Every 'theory of learning' that is worth considering incorporates an innateness hypothesis" (Chomsky, 1975a, p. 13) seems to refer to a general hypothesis. And "Suppose, we assign to the mind as an innate property, the general theory of language" (Chomsky, 1972, p. 88) refers to a general theory. (Of course, here Chomsky is not speaking about a hypothesis but a theory.) Finally, in 1995 he proposes that "something of this sort [P&P approach] is correct, but without trying to be very clear about the matter, since too little is understood to venture any strong hypotheses" (Chomsky, 1995, p. 6). This could indicate that earlier specific hypotheses had to be abandoned. But it does not explain why he would claim five years after this admission that there are "only *specific hypotheses* about the innate resources of the mind, in particular, its language faculty. General arguments against some unformulated 'innateness hypothesis' have no bearing on *actual hypotheses*

about innateness" (Chomsky, 2000b, p. 66) without providing an example of at least one such specific hypothesis.

Chomsky's subtle tactic of exploiting technicalities has allowed him to avoid addressing substantive criticism relating to the innate component of language and to maintain that his work has made steady progress in the absence of verifiable results. But it is dubious that it has helped to provide "appropriate concepts in a very sharp and clear form" (Chomsky, 2000b, p. 6).

Second, Chomsky has often exploited the fact that 'innateness' is an ill-defined term. He frequently uses near synonyms such as 'genetically determined' or 'biologically fixed' when he refers to the language faculty. Here are a few examples: "the genetically determined language faculty" (Chomsky, 1986a, p. 3), "the language faculty has an initial state, genetically determined" (Chomsky, 1995, p. 14), "[the language faculty is] part of our biological endowment" (Chomsky, 2000a, p. 4) and "a theory of the genetic component of the human language faculty" (Chomsky, 2007 a, p. 1097). However, he also freely uses the term 'innate'. Again I provide some examples: "innate mental functions that provide for the acquisition of [linguistic] competence" (Chomsky, 1967, p. 2), "an innate structure that is rich enough to account for [language acquisition]" (Chomsky, 1972, p. 79), "the innate endowment that serves to bridge the gap between experience and the knowledge obtained" (Chomsky, 1986a, p. xxvi), "a language acquisition device, an innate component of the human mind" (Chomsky, 1986, p. 3), and "innate resources of the mind" (Chomsky, 2000a, p. 66).

When Chomsky uses the term 'innate' he often does so without explaining

what *he* means by 'innate'. For example, it could be assumed that he equates 'innate' with both: 'genetically determined' and 'present at birth', when he writes, "UG is the theory of the 'initial state' of the language faculty prior to any linguistic experience" (Chomsky, 1986, pp. 4-5). However, in the same publication he writes "certain principles of UG are not available at early stages of language growth, obviously including the initial state S_0. … Genetically determined factors …are not to be identified with those operative at birth" (Chomsky, 1986a, p. 54). This shows that, if Chomsky equates 'innate' with 'genetically determined', then he cannot equate 'innate' with 'present at birth'.

There are more suggestions that for Chomsky 'innate' might be synonymous to 'genetically determined'. Chomsky writes that the "basic character [of the language organ] is an expression of the genes" (Chomsky, 2000a, p. 4), that the "initial state [of the language organ] is an expression of the genes" (Chomsky, 2002, p. 85), that "UG may be regarded as…genetically determined language faculty…an innate component of the human mind" (Chomsky, 1986a, p. 3) and that "certain properties are necessary properties of language…these properties are rooted in the genes" (Chomsky, 1983, p. 63).

Thus, it seems that for Chomsky a property is innate if it is genetically determined or an expression of the genes, part of our biological endowment, and in existence prior to relevant experience. I have discussed some of the problems that arise for such a view of 'innate' and, to my knowledge; Chomsky has not addressed these problems. Given that he purports to address important problems of "the biology of language" (Chomsky, 2000a, p. 5) it would be helpful if he could provide a precise

definition of the term 'innate' and then use the term consistently throughout his writings.

3.3. Innateness and Linguistic Nativism

Considering the important role that the notion of 'innateness' plays for linguistic nativists (e.g., Chomsky, 1957, 1966, 1975c, 1977, 1985; Fodor, 1975, 1981; Pinker, 1984; Lightfoot, 1999; Smith, 1999; Anderson & Lightfoot, 2002; McGilvray, 2002; Formigari, 2004; Cattell, 2006), it may come as a surprise that it is still not a well-defined term. As a consequence 'linguistic nativism' is not well-defined either. Usually linguistic nativism is contrasted with linguistic empiricism. In a first very broad approximation linguistic empiricists hold that all language-related knowledge is acquired from experience using general-purpose learning strategies (such as stimulus-response learning). Linguistic nativists claim that empiricists are wrong because language learning requires either additional domain-specific learning mechanisms or (at least some) innate knowledge of language or both.

Unfortunately, proponents of both sides often do not go beyond these broad definitions or explain innateness in terms of rationalism and vice versa. One of the many examples for such an uninformative definition is the following: "Rationalists ... are those who can plausibly deal with the issues that ordinary creativity and innateness pose for the study of mind. Empiricists are those who... do not, or cannot" (McGilvray, 2002, p. 9). Michael Stich observes that many proponents of innatism

"take the notion of innateness itself to be unproblematic. They explain it with a few near synonyms, 'inborn' or 'unlearned', or with a metaphor or an allegory" (Stich, 1975, p. 1). He cautions that it is crucial to specify exactly what we mean when we say that something is innate or is known innately. Stich was neither the first nor the last to encounter this unsatisfactory situation. Similar concerns have been expressed by Harman (1967), Wells (1968), Quine (1968), Hook (1968), Zimmerman (1968), Bunge & Ardila (1987), Sampson (1989), Pullum & Scholz (2002), Tomasello (2003; 2004), and Scholz & Pullum (2006). For example, Zimmerman (1968) points out that Chomsky wrongly implies that "empiricists deny that the ability to generalize or abstract is innate whereas what they deny is that the knowledge that is acquired by means of generalization etc. is innate" (p. 198). Roughly 40 years later Tomasello (2004) is dissatisfied with the use of 'innate' in linguistic debates. He observes that in many cases it is not evident

> …what type of innateness we are talking about, for example, Elman et al.'s (1996) architectural innateness or representational innateness... [and that evolutionary psychology] which proposes various innate cognitive modules including language, suffers from the same basic problem: everyone has a different list of innate cognitive modules, and there are no agreed upon methods for deciding among them. (Tomasello, 2004, p. 643)

In spite of these and many other requests for clarification the situation has not changed substantially. It is still common that critics of innatist's positions encounter the reply: "Your critique would be appropriate *if* innatism claimed what you say it does. But *that* is not what we claim". And, seemingly, for some defenders of innatism

virtually any mechanism that has been implicated in language acquisition can be innate:

> For the behaviorists, the innate constraints reside in the generalization gradients and response classes. For the connectionists, they reside in the features defining the units and the topology of the networks. For Chomskyans, they reside in categories, operations, and principles. For MacWhinney, they reside in the cues, items; alternatives pitted in competition, and categories whose absence constitutes 'indirect negative evidence.' Thus 'conservatism, item-based learning, indirect negative evidence, competition, cue construction, and monitoring' are not 'alternatives' to innate constraints on a learner's hypotheses, but claims about what those constraints are. (Pinker, 2004, p. 949-50).

Unfortunately, Pinker does not tell us any criteria for deciding which (if any) definition of 'innate' is correct and why.

Others have tried to be more specific. Elman (1999) suggests that innateness should be equated with constraints that can occur on several levels of (brain) development, given the expected inputs from the environment. He holds that constraints can occur "at the levels of *representations, architectures,* and *timing*" (p. 3, original emphasis). Representational innateness would occur if brain states, which underlie linguistic representations, were genetically pre-determined: "In brains, then, a claim for representational innateness is equivalent to saying that the genome somehow predetermines the synapses between the neurons" (Ibid.). Elman believes it is unlikely that this kind of innateness actually can be implicated in human language because it would require too much genetic specification. This view is shared even by critics of his account: "the DNA does not specify a point-by-point wiring diagram for the human brain" (Marcus, 2001, p.5).

The situation is somewhat different for the other two levels. Architectural constraints occur at the level of units (affecting specific properties of neurons such as firing threshold and the nature of pre-and postsynaptic changes), at the local level (affecting layering of brain tissue, packaging density of brain cells, degree and nature of interconnectivity) and at the global level (affecting how various subsystems of the brain are connected together). Although it seems that local and global constraints are at least to some degree genetically pre-determined, it is difficult to evaluate how much of the architectural differentiation occurring in adult brains is actually a result of environmental factors and development. Finally, chronotopic innateness occurs when the outcomes are "constrained through the timing of events in the developmental process... In brains, timing is sometimes under direct genetic control, but the control of timing may also be highly indirect and the result of multiple interactions" (Ibid., p. 5). In these cases we also have a highly complex interaction between genetic and environmental factors. This account is similar to the one proposed by Elman et al. (1996) and it has been critiqued by Marcus (2001) as essentially unsatisfactory.

Marcus calls this account "a sort of stripped- down nativism in which 'architectural' aspects of brain organization are innate, but 'representations' are not ...[and he argues] that that the stripped- down nativism of Elman et al probably relies too much on experience" (p. 7). According to Marcus, Elman puts the burden of brain organization primarily on a basic form of 'learning'. He holds that the effect of the genetically predetermined organization of neurons is negligible in comparison to the impact of environmentally determined experience. Essentially Marcus rejects

Elman's account because he believes that for Elman developmental flexibility of brain organization entails that massive 'learning' from the environmental inputs takes place. However, it has been shown that many developmental changes take place that cannot sensibly be described as learning (e.g., differentiation of eye, stomach or kidney cells).

Second, Marcus holds that Elman's account mistakenly equates innatism with the idea of the DNA as a blueprint that specifies in advance precise details of brain development. But defenders of innatism acknowledge that such an account is not plausible because the human genome does not contain enough information to specify the location of neurons or synapses and because we know that extensive rewiring of brain regions takes place in response to injury.

Finally, Marcus alleges Elman's account rests too heavily on learning and neural activity and attributes virtually all detailed brain organization to neural activity. This is a problem because, especially in the earliest stages of brain development, the contribution of neural activity is minimal. Thus it is implausible that brain organization arises solely as result of learning (Ibid., p. 374).

To overcome these limitations, Marcus suggests, we need to pay more attention to the 'toolkit of developmental biology', to explore the array of techniques for self-organization of incredibly complex organisms. Marcus proposes modeling the innate brain structures by building artificial neural networks "that will integrate ideas about nodes and connections with some of the basic principles of developmental biology" (Ibid., p. 375). These include cell -division, -migration, and -death, gene expression, cell-to-cell communication, and cascades (triggering of gene expressions

through proteins or other genes). Marcus argues that his top-down approach allows him to model "how brains with particular properties could be assembled by genetic-like processes" (Ibid., p. 376). This in turn would show us something about innate structuring of brains.

While the 'naturalness' of Marcus' models is an attractive feature, it is not clear how closely they can resemble the processes that occur in actual brains during development. It takes several years for a human brain to mature, and the intricacy and complexity of the connections that emerge during this process are well beyond the reach of existing models. In addition Marcus is convinced that brains have particular innate properties that underwrite rule following. This conviction could affect which of the current models he considers to be most 'brain-like'. The evidence he cites is not conclusive, and a more charitable interpretation of the Elman models is possible.

The foregoing shows that there is considerable debate between language acquisition researchers about an adequate model of innateness. One of the promises of Chomskyan linguistics was to bring precision to linguistics and to place linguistic research squarely in the natural sciences (e.g., Chomsky 1957, 1975d, 1980, 1985). If, as McGilvray claims, Chomskyan linguistics has made steady progress over the past five decades we should hope that it has contributed to clarifying the definitional issues discussed in the previous section. To evaluate whether or not this is the case I will now provide a short overview of the evolution of Chomsky's linguistics, focusing on two questions: 1. Has Chomskyan linguistics contributed to better definitions of key terms such as 'innateness'? 2. Has Chomskyan linguistics provided insights that allow us to situate language better within the natural sciences? I will

argue that the answer to both questions is 'no'.

3.4. Chomskyan Linguistics - A Brief History

3.4.1. Chomskyan Linguistics as Natural Science

In the following sections I will highlight some of the important milestones in the evolution of Chomskyan linguistics. Because throughout his career Chomsky has put much emphasis on treating linguistics as part of the natural sciences, his definition of scientific theory is relevant. In his early work this definition is uncontroversial.

> Any scientific theory is based on a certain finite set of observations and, by establishing general laws stated in terms of certain hypothetical constructs, it attempts to account for these observations, to show how they are interrelated, and to predict an indefinite number of new phenomena. A mathematical theory has the additional property that predictions follow rigorously from the body of theory. (Chomsky, 1956, p.113)

According to this definition, the linguist attempts to construct a grammar for the language she investigates. Her task is well defined: "The grammar of a language can be viewed as a theory of the structure of this language … A properly formulated grammar should determine unambiguously the set of grammatical sentences" (Ibid.). Chomsky considered it to be essential for scientific progress to develop precisely constructed models and criticizes those who do not share this view:

I think that some of those linguists who have questioned the value of precise and technical development of linguistic theory may have failed to recognize the productive potential in the method of rigorously stating a proposed theory and applying it strictly to linguistic material with no attempt to avoid unacceptable conclusions by *ad hoc* adjustments or loose formulation. Chomsky, 1957, p. 5)

Here Chomsky reminds fellow linguists that precise theories are not merely one of many tools for scientific inquiry but form the foundation of such inquiry. Deviating from rigorous models will lead to untestable or unfalsifiable hypotheses and hinder genuine scientific progress. This early commitment to scientific rigour has been acknowledged even by linguists critical of Chomsky's view: "Chomsky's exposition of how in principle the syntax of a language can be brought within the purview of scientific linguistic description is a great positive contribution" (Sampson, 1980, p. 134).

Over the years Chomsky has gravitated away from this well-defined view of theory construction. I will trace these changes here. At first the changes were not significant and the aim still seemed to be a well-defined empirical project:

The study of universal grammar…is a study of the nature of human intellectual capacities. It tries to formulate necessary and sufficient conditions that a system must meet to qualify as potential human language, conditions that are…rooted in the human 'language capacity', and thus constitute the innate organization that determines what counts as linguistic experience and what knowledge of language arises on the basis of this experience. Universal grammar then constitutes an explanatory theory… Chomsky, 1968, p. 24)

Formulating a set of necessary and sufficient conditions for human languages that are rooted in our biology would have resulted in testable hypotheses.

135

Looking back at his early work, Chomsky recalls that the "central objective was to abstract general principles from the complex rule systems devised for particular languages" (Chomsky, 1995, p. 5). Empirical work has repeatedly revealed redundant principles that were "wrongly formulated and [had to be] replaced by nonredundant ones" (Ibid.). Thus the early work of Chomsky clearly fell within the generally accepted realm of science: he formulated testable hypotheses, and, when data disconfirmed specific predictions, the hypothesis was adjusted accordingly.

By the 1970s Chomsky's conceptualist approach to linguistics had virtually replaced the Bloomfeldian structuralist/descriptivist approach and had become very influential in Northamerican linguistics. Chomsky's efforts to turn linguistics into a natural science were considered successful and his work was highly regarded even by those who disagreed with him about important details of his proposals:

> "I should like to dedicate this study to Noam Chomsky. The present work is highly critical of some of his recent grammatical proposals. But… none of this kind of work would ever have been possible without the many fundamental and groundbreaking insights and the radical reorientation of the goals and methods in linguistic inquiry which he has played such an enormous role in bringing about" (Postal, 1974b, p. iii)

Postal states here that he has some factual disagreements with Chomsky. But he also clearly acknowledges the positive impact of Chomsky's groundbreaking work on linguistics. Similarly, another linguist who disagreed with Chomsky on several fundamental issues, stated, "…[my] book will strongly support Chomsky's novel way of looking at language" (Sampson, 1975, p. 9). Seemingly, until this time it appeared that Chomsky's work was contributing to genuine progress in linguistic

136

research. He made innovative proposals and his theories seemed to be formulated precise enough that they could be disconfirmed by empirical evidence.

Yet, by the late 1980s it became clear that the "radical break from the rich tradition of thousands of years of linguistic inquiry" (Ibid.) that had motivated the work of the 1960s and 1970s was in need of further substantial revisions. The outline of the research projects of this time still seemed to fall within the confines of scientific inquiry

> The theory of UG must meet two obvious conditions. On the one hand it must be compatible with the diversity of existing (indeed, possible) grammars. At the same time UG must be sufficiently constrained and restrictive in the options it permits so as to account for the fact that each of these grammars develops in the mind on the basis of quite limited evidence…. What we expect to find, then, is a highly structured theory of UG based on a number of fundamental principles that sharply restrict the class of attainable grammars (Chomsky, 1986a, p. 3).

Here Chomsky focuses on searching for a highly structured theory of Universal Grammar that can explain the diversity of existing languages but also constrain the search to languages alone. This research project generated a wealth of publications during the 1980s. Yet, as we shall see in section 3.5., attempts to fill in the necessary details failed and were eventually abandoned. While the lack of verifiable results led eventually to another major revision of the theory Chomsky's also began to express suspicion towards data gathering and scientific research:

> I certainly feel that explanation is much more important than gross coverage of data…gross coverage of data is much too easily obtained, in too many different ways. There are millions of ways in which you can give a kind of rough characterization of a lot of data, and therefore you

don't learn anything about the principles at all (Chomsky, 1981, p. 83)

Chomsky is correct; one can always interpret data in numerous ways. Yet, because he does not qualify his remark, it gives the reader the impression that aimless data-gathering is common in empirical research and results in work that will not contribute to our understanding of language. It would have been better to provide specific examples that can support the suggestion that this is a fair criticism of actual research. As we shall see in chapter 5, scientists who actually gather data and analyze them carefully discover at times facts that have not been predicted by their hypotheses. Thus unless Chomsky has an infallible *a priori* method that allows him to select the correct explanations without need for data-gathering, his blanket dismissal of such empirical work seems unjustified. Yet, he continues to express a dismissive attitude towards data gathering:

> The phenomena that surround us in the real world of ordinary experience are generally not understood in any clear manner; they are too complex, they involve too many interacting factors, and our understanding is too limited… Most data fail to help us attain insight into underlying principles and structure and are therefore uninteresting for the purpose of attaining rational understanding. (Chomsky, 1991, p. 42)

Chomsky is correct to assert that all scientific inquiry needs to focus on a well-defined set of data and that many observations we can make are irrelevant to the phenomenon we try to study. But again, no specific work is cited, and Chomsky does not demonstrate how work that has been completed has failed to produce relevant data. Instead, he seems to imply that most of the data gathered by

language-acquisition researchers testing children and computational models of language acquisition are 'uninteresting'. Seemingly, he made this assessment based on his rationalist commitments. This was a break from his earlier assertions that the questions of linguistics are empirical questions that will be answered through empirical inquiry (e.g., Chomsky, 1966, 1975a, 1980). Going hand in hand with this changing attitude was a terminological revolution. Instead of developing theories, defending hypotheses or even principles, Chomsky was now referring to "rather vague but yet intelligible guidelines" (Chomsky, 1986a, p. 339). And when asked for verifiable results he replied "almost everything is subject to question, especially if you look at it from a minimalist perspective" (Chomsky, 2002, p. 70). During this period his attitude towards commonly accepted procedures apparently changed as well.

In the scientific community it is widely accepted that researchers need to develop "precise, testable, computationally tractable models of language" (Evans & Levinson, 2009, p. 475) and, should these models make repeatedly wrong predictions, be prepared to rethink the conceptual foundations of the models. What needs to be avoided is the construction of theories that can be maintained regardless of empirical findings because it is impossible to falsify their predictions. Yet, the more recent formulations of Chomsky's theories (e.g., the minimalist program) have become so vague that it would be difficult to imagine how they could be falsified (for discussion see section 3.5; and Pullum, 1996; Newmeyer, 2003; Seuren, 2004; Culicover & Jackendoff, 2005; Tomasello, 2009) or could be seen as a continuation of earlier work: "With the M[inimalist] P[rogram] we see a shift to a deep

skepticism about formal devices of traditional syntactic theory that are not in some sense reducible to 'virtual conceptual necessity'. Such a perspective thus explicitly rules out precisely the major theoretical achievements of the past. *All of them*" (Culicover, 1999, p.138, emphasis added). Yet Chomsky continues to call his own approach scientific, and he also acknowledges the importance of empirical research:

> The problems [of linguistic inquiry] are nevertheless empirical, and we can hope to resolve them by learning more about the language faculty and the systems with which it interacts. We proceed in the only possible way: by making tentative assumptions about the external systems and proceeding from there. (Chomsky, 1995, p. 222-3)

Here Chomsky clearly accepts that empirical research will bear on the answers to the questions in linguists. Moreover, he places his own work within the scientific framework. Yet, at the pre-dawn of the new millennium Chomsky expresses again profound distrust in scientific work conducted by others:

> Science is a very strange activity. It only works for simple problems… The idea that deep scientific analysis tells you something about problems of human beings…is mostly pretence in my opinion, self-serving pretence which is itself a technique of domination and exploitation (Chomsky, 2000b, p. 2)

Again, this general statement is not supported by any detailed and targeted critique of specific work. Here, as elsewhere, Chomsky expresses a dismissive attitude towards the work of the scientific community. Nevertheless, he still considers his own research science: "…the area of human language…is one of the very few areas of complex human functioning [where the approach of natural sciences seems to

get somewhere]" (Ibid.). And he continues to claim that his work aims at "embedding linguistics more fully within biology, including evolutionary biology" (Chomsky, 2007 a, p. 1097).

In the same year he writes that his linguistics is an "empirical discipline … in which theories are put forth *tentatively as 'best guesses'*" (Chomsky, 2007, p. 1100, my emphasis). This appears to be a correct description for a science that newly emerges and has not gathered any data. But, according to Chomsky, his own work is to some degree a "rediscovery of much that was well understood in [the Cartesian] period" (Chomsky, 1966, p. 1), and his school has continually produced results since the 1950s. Thus, one would expect that he is past the 'guess-work' stage and has formulated 'appropriate concepts in a very sharp and clear form…[and given] a precise account of the computational principles that generate the expressions of a language" (Chomsky, 2000a, p. 6). Such an account is still wanting.

Recently one supporter of Chomsky has offered an explanation for the discrepancy between the expectations generated by Chomsky's rhetoric of rigour and precision and the results of his 'tentative guesses'. He suggests that Chomsky is not bound by the requirements of normal scientific procedures: "Can [Chomsky's] goal be pursued by 'normal scientific procedure'? Let us remember we are talking about someone who tries to reinvent the field every time he sits down to write. Why should we expect Chomsky to follow normal scientific practice…?" (Fiengo, 2006, p. 471). Given that this comment was made in response to Seuren's (2004) critique of Chomsky's work it is relevant what, according to Seuren, scientific practice is. If they are overly restrictive Fiengo might be justified to reject them. However, Seuren's

acceptability requirements for scientific work are quite modest. Scientific work "should satisfy certain....elementary criteria of common rationality: clarity of expression; faithfulness to the relevant facts....; acceptance of whatever can be concluded on the basis of these facts; and a mandatory preference for ... the theory that explains more facts with less machinery than rival theories" (Seuren, 2004, p. 10). It would seem if Fiengo and Chomsky reject these criteria, it is legitimate to ask if the activity they engage in can be called scientific.

In section 3.4.3. I will discuss some of the details of Chomsky's work and evaluate whether or not it has contributed to a better understanding of linguistic phenomena in general and of language acquisition in particular. Before doing this I want to introduce some targeted criticism regarding the claim that Chomsky has succeeded in situating linguistics firmly within the natural sciences. I will argue that these challenges have not been addressed by Chomsky's work.

3.4. 2. Critique of the Chomskyan Science

The claim that Chomsky's research situates linguistics within the biological sciences has been challenged by many critics (e.g., Pullum, 1996; Seuren, 1998, 2004; Sampson, 2001; Postal, 2003a, 2004, 2009; Edelman & Waterfall, 2007). Especially, his research methodology has become the target of explicit criticism:

> Chomsky's own work in linguistics… nowhere resembles biological research…I am unaware of any reports by Chomsky of his use of x-ray machines, microscopes, tissue samples, and so on. So in total contrast to actual biological science, in four decades he has not specified *a single*

physical property of any linguistic object. (Postal, 2009, p. 113, original emphasis)

Before addressing Postal's criticism I would like to draw attention to the fact that even an experienced linguist like Postal does not seem to have a clear conception of the requirements of biolinguistic research. Maybe a first time visitor of a science fair would expect to encounter x-ray machines and microscopes. But these hardly seem to be instruments one would use to investigate the brain in search of language-specific structures. This seeming lack of knowledge is curious given that Chomsky claims that his work is investigating "the biology of language" (Chomsky, 2000a, p. 5). Thus we should expect that a considerable part of Chomskyan linguistics would be conducted as biological research. And, as customary, the specific scientific instruments and procedures used for this research would be documented in detail in linguistic publications. So even a linguist who does not work himself in the lab would be at least somewhat familiar with the experimental set-up of the biological research. That this apparently not the case might indicate that Postal's suggestion (that no biological research is conducted by Chomskyan linguists) is correct.

We know that Chomsky himself dislikes experimental work. He is famous for the comment "I hate experiments" (cf., Mehta, 1971, p. 209). Now refraining from experimental work when one does not posses an aptitude for it is actually Cartesian in spirit: "…there are some experiments which only a few people are capable of performing, and when they are performed badly, the result is often the very opposite of what it ought to be" (CSMK III, p. 90). For this reason, it is usual

practice in the well-established sciences that some people specialize in theorizing and others in experiment. Thus, the fact that Chomsky himself does not conduct any experimental work *alone* does not imply that his work has not contributed to biology in particular and the natural sciences in general. Developing an adequate framework for linguistic research or clarifying problematic concepts (e.g., innateness) certainly would be a valuable contribution.

Nevertheless, a substantial part of work in the natural sciences consists in investigating the physical properties of the objects that have been postulated by the theory. This has been also acknowledged by defenders of Chomsky: "...the psychological reality, if any, of transformational generative linguistics can only be shown through proper laboratory experiments administered by psychologists" (Leiber, 1988, p. 309). Chomsky has insisted for decades that I-language, the proper object of linguistic investigation, is part of the human brain (e.g., Chomsky, 1980a, 1985, 1986b, 1995). For this reason Postal is justified to ask for progress in actual brain research based on the Chomskyan paradigm. Attempting to trace down how Chomsky's work has affected neuro-science and brain research would exceed the scope of this dissertation. But such impact needs to be documented by defenders of Chomskyan science.

Postal is not the only one who suggests that Chomsky's methodology is deeply flawed. Others have pointed out that Chomsky describes work that can and should be done in order to confirm his hypotheses in a manner that creates the impression it has already been completed: "a program for research that might one day be undertaken should not be confused with one already completed. That is what

has been happening up to now" (Pullum & Scholz, 2002, p. 11). Pullum and Scholz refer to the fact that many of the 'results' that are used in Chomskyan theorizing have never been confirmed by empirical research. However, these 'results' have been used as justification in further theorizing. For example Chomsky (1980) claims that native speakers share intuitions about how to interpret grammatical structures (sentences or utterances). This claim in turn has been used as confirming evidence for the claim that we have an innate Universal Grammar (UG) module or language acquisition device (LAD) that is 'uniform across the species'. This LAD determines how we interpret sentences and explains that the interpretation is the same for all competent native speakers (Chomsky, 1980). Yet, there are no studies cited that have ever tested this postulate.

In fact one can easily find evidence challenging this postulate. As a non-native speaker of English, I do not always trust my intuitions, and when I encounter examples that strike me as ambiguous, I ask native speakers about the correct interpretation. It turns out that native speakers do not always agree. For example the following sentence is ambiguous:

Does the inability of other species to develop language of the human type derive from their lack of a specific quality of intelligence rather than from a mere limitation in a common intelligence, as Descartes thought? (Chomsky, 1975a, p. 40)

It is not clear if the clause 'as Descartes thought' refers to 'their lack of a specific quality of intelligence' or to 'a mere limitation of common intelligence' or to both. From the 39 people I surveyed only 4 thought the sentence was not

ambiguous (2 of these subsequently changed their mind). From the 34 people who thought the sentence was ambiguous 12 claimed it is genuinely ambiguous and cannot be interpreted. From the 23 people who thought that even though the sentence is ambiguous, we can be fairly sure what it means, 8 thought it was likely that 'as Descartes thought' refers to 'their lack of a specific quality of intelligence', 12 thought to 'a mere limitation of common intelligence' and 3 thought 'as Descartes thought' refers to both: 'their lack of a specific quality of intelligence' and 'a mere limitation of common intelligence'. It is clear that this informal survey does not provide sufficient evidence against Chomsky's claim that intuitions regarding interpretation are uniform across native speakers. However, the ease with which I found a potential counter example suggests that we would need to design and perform specific experiments to either confirm or disconfirm Chomsky's claim.

When we turn to the linguistic literature we easily find more anecdotal evidence that the Chomskyan dictum is in need of empirical confirmation. An increasing number of publications show that linguistics intuitions are neither uniform nor reliable. Because of space limitations I mention only a few examples here. In 1962 Chomsky claimed at a conference that 'the verb perform cannot be used with mass-word objects: one can perform a *task*, but one cannot perform *labor*. This generalization was almost immediately proven wrong when a speaker observed that 'to perform magic' is grammatical (for details see Pullum, 1972, pp. 57–58).

In another case, Bauer (1990) describes a generalization about the subcategorizational behaviour of prefixed words that has been accepted and widely repeated in spite of the fact that counterexamples can easily be found. Carlson and

Roeper (1981) claim that "addition of prefixes to verbs rules out non-nominal complements" (Carlson & Roeper, 1981, p. 123). Bauer lists numerous counter examples to this alleged general rule:

(a) Pip understood hat I could not come.
(b) Lee disclosed that no-one had been informed.
(c) Sandy predicted that Bush would win the election.
(d) She has miscalculated badly.
(e) He prefers to suffer than to ask for help.
(f) Robin had predetermined to allow her access.
(g) The minister reiterated that no such plans existed.
(h) Pat undertook to pay the money back within six months.
(i) Sam has a tendency to overindulge in chocolate.
(j) Fate had preordained that we should fail.
(k) Chris reaffirmed that everything possible was being done. (Bauer, 1990, pp. 25-26)

Bauer demonstrated here that we cannot trust blindly the intuitions of competent speakers. There is no doubt that Carlsen and Roeper, both distinguished linguists, are competent speakers. Yet, their generalization turned out to be wrong. Similar examples are discussed in Seuren (1998), MacWhinney (2004), Sampson (2001, 2005), Pullum (2007), and Leon (2010).

In spite of mounting evidence about the diversity of native speaker intuitions Chomskyans continue to rely on these intuitions and consider the use of introspection not as a supplementary source of linguistic data but give it "an authority which is denied to observation" (Sampson, 1980, p. 153). This methodology has been criticized for decades. As discussed above, empirical research has provided a wealth of counter-examples to the claim that speakers of one language community universally hold strong intuitions about grammaticality or ungrammaticality of sentences. Yet Chomskyans continue to rely mainly on native

speaker intuitions (e.g., McGilvray, 2005, 2009; Hornstein & Boeckx, 2009). In my opinion, it would be desirable to engage constructively with well-known criticism and develop a productive dialogue with non-nativist researchers.

For the reasons discussed here Chomsky's attitude towards natural sciences and towards the gathering and evaluation of data has come under severe criticism. Sampson observes that there is "a tendency for Chomskyan linguists to abandon the principle that science is cumulative" (Sampson, 1980, p. 159). He claims that this tendency is a consequence of the rationalist commitments of Chomsky's work:

> The rationalist …thinks of the individual as 'inheriting' true knowledge in the genetic sense…the thought of previous generations is redundant insofar as it is correct, and merely misleading where it is wrong. …the consequence of this is that much research by members of the Chomskyan school, even when it is not vitiated by reliance on fallacious introspective judgments, consists of time wasting rediscovery of facts that had long been common knowledge outside the Chomskyan camp. (Sampson, 1980, p.159-60)

From this perspective the rationalist approach championed by Chomskyans is a hindrance for genuine scientific progress, and, unsurprisingly, Sampson claims that such progress has been lacking and that the only valuable contribution of Chomskyan linguistics "was already stated, about as adequately as it ever has been since, in Chomsky's first book [Syntactic Structures]" (Sampson, 1980, p. 164). This contribution was the recognition that "all languages can be generated by constituency grammars" (Ibid.). According to Sampson the flurry of research activity of the following decades has added nothing to this original insight. More recent commentaries share this unfavourable verdict (e.g., Seuren, 1998, 2004; Postal, 2004; Boden, 2006; Edelman & Waterfall, 2007). In section 3.5 I will show

that these criticisms are not just tendentious rhetoric but highlight Chomsky's inability to produce verifiable results that can be analyzed by researchers outside of his own school.

Most researchers will probably agree with the observation that "the field [of linguistics has been] rapidly changing under the impact of new empirical material and theoretical ideas" (Chomsky, 1995, p. 10). However, there is little consensus among present-day linguists and philosophers that the repeated "major conceptual shifts" (Chomsky, 1985, p. 7) and "radical departure[s] from [earlier views]" (Chomsky, 2000a, p. 14) constitute genuine progress. What is at stake is not the claim that Chomskyan linguistics "is a living field, in which sane investigators are constantly changing their minds in the light of new empirical discoveries and theoretical insights, a regular occurrence" (Chomsky, 2007 a, p. 1098). Rather, researchers are concerned about measurable indications of genuine progress that has been produced by the Chomskyan enterprise.

In the following section I will introduce milestones of the evolving Chomskyan theorizing and inquire whether they withstand the criticisms discussed here. The specific questions that need to be answered are (i) has Chomskyan theorizing added rigour and clarity to linguistics, (ii) does Chomskyan science provide a coherent framework for linguistic research, (iii) has the research conducted by Chomskyans contributed to a better understanding of the acquisition and use of language, and (iv) does Chomskyan science depart from 'common scientific practice', and, if so, have these departures been acknowledged and justified?

3.4.3. From Standard Theory to Minimalism

Early in his linguistic career Chomsky focused on syntax and grammar (e.g., Chomsky 1956, 1957, 1959, 1965a, 1966, 1968). Like Descartes, he believed that language distinguishes humans from other animals and proposed that the unbounded infinity of language is in need of explanation. But while for Descartes our ability to *understand* the content of ideas that goes beyond our sensory experience (God, abstraction, mathematical concepts, etc.) was in need of an explanation, the early Chomsky seemed not interested in semantics. For him it was not the content of our ideas but the rules for the combination of these ideas into a potentially infinite set of sentences that required an explanation.

Chomsky was able to overcome limitations of linguistic approaches that had dominated the field until the 1950s (e.g., Bloomfield's structuralism, Skinner's Behaviourism). The structuralist theories of taxonomic grammars limited grammatical analysis to segmenting and classifying utterances. The goal was to obtain a comprehensive catalogue of a corpus (a given natural language). Because structuralist grammars are limited to analyzing actual utterances they can ultimately not account for the generative aspect of human language. This became apparent after Zellig Harris (1951) had developed a theory of transformational grammar (for details of this development see Katz, 1981, pp. 24-34). The implications of this theory required a change of the ontological framework of linguistics and Chomsky's work provided a way to achieve this change:

Chomsky's real innovation was to provide a new way of interpreting the formal apparatus of transformational grammar, which both recast the form of transformational grammar from the taxonomic to the generative and the subject matter of transformational grammar from the external acoustic effects of speech to its internal psychological causes. (Katz, 1981, p. 24)

Some of the key claims of Chomsky's early work were (i) that human languages have syntactic universals, (ii) a grammar defines the class of grammatical sentences, and (iii) the universals define a range of humanly-possible grammars (and by implication rule out any logically-possible grammar not contained within that range). Sentences are composed of constituents (verb phrases, noun phrases etc.), and these constituents can be either simples (e.g. a single noun for a noun phrase) or composed of constituents themselves. A grammar provides a finite set of rules for combining constituents into larger units (sentences) and can generate an infinitely large set of sentences (language). Languages, which are generated by a constituency grammar (also called context-free phrase-structure grammar), are a well-defined subset of the set of all possible languages. Chomsky claimed that human languages couldn't be generated by constituency grammars alone and proposed that an additional series of transformational rules is needed to generate all grammatical sentences of human languages:

I will consider a language to be a set (finite or infinite) of sentences, each finite in length and constructed out of a finite set of elements. All natural languages are languages in this sense. …The fundamental aim of linguistic analysis of a language L is to separate the *grammatical* sequences, which are sentences of L from the *ungrammatical* sequences,

which are not sentences of L, and to study the structure of the grammatical sequences. The grammar of L will thus be a device that generates all of the grammatical sequences of L and none of the ungrammatical ones. (Chomsky, 1957, p. 13)

This early work contributed to clarifying important conceptual issues and sketched the outlines of a novel scientific framework for linguistics. The great potential of this work was recognized quickly. When Morris Halle and Chomsky established a linguistics programme at MIT it "immediately attracted a number of gifted scholars… All were eventually named to MIT faculty - Lees and Postal in linguistics, Fodor and Katz in philosophy (Barsky, 1997, p. 101). Some thirty years later one of these gifted scholars vividly recalled the initial appeal Chomsky's work had "I was very impressed, first with the power of his thought, but also [by Chomsky's work that]... was based on an entirely different way of thinking from anything I had come into contact with before" (Postal quoted in Harris, 1993, p. 102).

The value of Chomsky's initial contribution has been acknowledged by others linguists as well; both on a personal level: "My indebtedness to the work of Professor Noam Chomsky of MIT, Cambridge, Mass. will be evident on almost every page of this book" (Seuren, 2010/1968, p. x) and as contribution to the field: "Chomsky's exposition of how in principle the syntax of a language can be brought within the purview of scientific linguistic description is a great positive contribution to the discipline" (Sampson, 1980, p. 134). Similar praise for Chomsky's early contribution is expressed here: "Chomsky [1957]… issued what is still the finest and most cogent defense of the formalist position that linguistics has ever had...Never was it put better before or since" (Pullum, 1991, p. 50). It has been also acknowledged that Chomsky's

work had impact beyond linguistics: "[Chomsky's work can help] developing psychology to incorporate the sophisticated formal insights that generative grammar has produced" (Pullum, 1972, p. 64). Others add that Chomsky's early work also made a valuable contribution to computer modeling: "Chomsky's influence on cognitive science was beneficial in many ways…He offered a vision of theoretical rigor that inspired linguists and non-linguists alike. And…his work encouraged others to attempt the computer modeling of mind" (Boden, 2006, p. 591). In addition it is widely recognized that Chomsky contributed greatly to overcoming the behavourist approach to linguistics (e.g., Sampson, 1980; Botha, 1989; Seuren, 1998; Boden, 2006).

The focus on syntax and the search for formal descriptions of transformational rules continued to motivate Chomsky's research. Throughout the 1960s he continued to describe language as "a set (finite or infinite) of sentences, each finite in length and constructed of a finite set of elements" (Chomsky, 1965a, p. 13), and he held that not the elements but the rules for their combination form the essence of language. He proposed that it is possible to gain an understanding of grammar without paying any attention to semantics: "grammar is autonomous and independent of meaning, and ... probabilistic models [based on meaning] give no particular insight into some of the basic problems of syntactic structure" (Ibid., p. 17). Chomsky suggested that finite state grammars (those that begin transformations according to specified rules at an initial state and, after running through a sequence of intermediate states, end in a final state) produce finite state languages. Natural languages such as English are not finite state languages. This means that no finite state grammar "will produce all and only

the grammatical sentences of English" (Ibid, p.21). This point was crucial for Chomsky. Because no finite state grammar can "account for the ability of a speaker of English to produce and understand new utterances, while he rejects other new sequences as not belonging to the language" (Ibid., p.23), his ability to interpret sentences he has not heard before cannot be explained based on the input that a language learner receives. Or so Chomsky argued. We will return to this point in section 3. 5.

As we have seen, syntactocentricsm was essential for Chomsky's early accounts of language. He emphasized that language has two syntactic levels: deep structure and surface structure. While much work in previous linguistics had focused on surface structure, Chomsky (1965a, 1968) proposed that, in order to fully understand grammars, we need to study deep structure. According to Chomsky (1965a) grammars have a 'tripartite structure' consisting of

- a sequence of rules from which the phrase structure can be reconstructed
- a sequence of morphophonemic rules that convert strings of morphemes into strings of phonemes
- a sequence of transformational rules transforming strings to which phrase structure applies into strings to which the morphophonemic rules can apply.

On this view phrase structure and morphophonemic rules are elementary because they only require knowledge of the shape of the string to which they apply.

By contrast transformational rules are complex because they require not only knowledge of the shape of the strings but also knowledge of some of the history of the derivation of these strings (Chomsky, 1965a, p. 107).

To accommodate some of the findings of his early work Chomsky proposed an extended standard theory (EST) in the late 1960s and continued refining it until 1973. This theory assumes that grammar is based on a rule system of the following general organization: simple phrase structure rules generate an infinite class of D-structures (deep structures) that express semantically relevant grammatical functions and relations; transformational rules convert D-structures to S-structures (surface structures). Phonological and other rules convert S-structures to phonetic representations with surface phrasal categories, and, independently, rules of the LF (logical form) component convert S-structures to representations in LF, where scope and other properties are directly represented (Chomsky, 1965a, p. 67-8).

In the 1980s Chomsky introduces a conceptual shift. While he initially conceived of languages as potentially infinite sets of sentences, he now sees languages as physical parts of biological brains. He explicitly rejects an analogy between language and arithmetic:

> …it has been suggested that knowledge of language should be understood as the analogy of knowledge of arithmetic, arithmetic being taken to be an abstract 'Platonic' entity that exists apart from mental structures…The analogy [is] quite unpersuasive. In the case of arithmetic, there is at least a certain initial plausibility to a Platonistic view insofar as the truths of arithmetic are … independent of any facts of individual psychology. In the case of language however, … there is *no plausibility* to the idea that apart from the truth of grammar concerning the I-language and the truth of UG concerning S_0 there is an additional domain of fact about P-languages, independent of any psychological states of individuals. (Chomsky, 1985, p. 27, emphasis added).

We can see that for Chomsky there is a fundamental difference between arithmetic and language. Only in the case of arithmetic it may make sense to speak about the study of abstract objects that exist independently of human brains. In the case of language no such objects (P-languages) could exist independently of human brains. It is at times not easy to discern what Chomsky means because in the above quote he seems to refer to two entities when he talks about the truth about I-language *and* the truth about UG. However, we need to understand that Chomsky uses the terms 'Language acquisition device', 'Universal Grammar', 'language faculty' and 'language' interchangeably as referring to the same object (see Chomsky (2000a) and discussion below). One attempt at clarifying his ontological realism explains that:

> …a mentally represented grammar and UG are real objects, part of the physical world, where we understand mental states and representations to be physically encoded in some manner. Statements about particular grammars or about UG are true or false statements about steady states attained or the initial state (assumed fixed for the species), each of which is a definite real-world object, situated in space-time and entering into causal relations. (Chomsky, 1983, p. 156-57)

According to Postal (2009) the attempted conceptual move from 'language' as an abstract set of sentences to 'I-language' as a physical part of human brains is highly problematic. Postal claims this move has never been fully completed (for similar criticisms see Cummins & Harnish, 1980; Langendoen & Postal, 1984; George, 1987; Katz & Postal, 1991; Katz, 1985, 1996; Postal, 2003a). Instead, Chomsky continues to treat language as both: a set of sentences and a biological object (e.g., Chomsky 1986b, 1995, 2000a, 2007b, 2009a, 2010b). But this treatment

blurs the distinction between the object of linguistic study (sentences of a language and their logical relations) and the object of physiological/neurological study (brain structures involved in generating the sentences linguists can analyze). The incomplete conceptual shift of the early 1980s commits Chomsky to the untenable view that languages are both finite (as parts of finite human brains) and potentially infinite (as grammatical strings of words).

The problems Postal highlights arise even if we do not assume that languages *are* literally infinite sets of sentences but allow that languages consist of *potentially* infinitely many sentences or that sentences are of finite but unrestricted length. This is the case because for Chomsky language is literally located in the human brain, which is finite and of definite size. Yet, the rules of generative grammar allow for sentences of unrestricted length. But sentences that actually could be stored in the brain of a human have, by necessity, a finite length. For example, Chomskyan grammar can generate the sentence "I own a blue slab and a brown slab and a red slab and a blue and brown slab and another blue slab and…". This sentence is of unrestricted length. It cannot be stored in the human brain, because at any given brain-storage capacity we could add at least one more "and another blue slab" to the sentence. Thus, only a finite segment of the complete sentence can be so stored. This is not an inconsistency problem for all physicalist/naturalist views of language. It plagues only those who, with Chomsky, hold that language is both: a (potentially or actually infinite) set of sentences of potentially unrestricted length *and* a part of the biological brain.

Chomsky's clearly is committed to the view that languages are parts of our brains:

…we may think of a person's knowledge of a particular language as a state of the mind, realized in some arrangement of physical mechanisms. We abstract the I-language as 'what is known' by a person in this state of knowledge. This *finite system*, the I-language, *is what the linguist's generative grammar attempts to characterize* (Chomsky, 1986a, p. 40, emphasis added)

Here Chomsky clearly states that a linguist's grammar characterizes a finite object. Earlier in the same publication Chomsky explains that traditionally generative grammar has been understood as "characterization of [the] infinite set of [sentences or sentence-meaning pairs]" (Ibid., p. 30) and claims that "the alleged problems concerning the concepts grammar and knowledge of grammar…[may be]…traced to unfortunate terminological choices which reinforced inappropriate analogies to the formal sciences" (Ibid., p. 29). Next, I will discuss an attempt by Chomsky to clarify the issue and show that it is not successful.

During the 1980s Chomsky introduces, what he considers to be an essential distinction: (externalized) E-language and (internalized) I-language. He suggests that only the I-language is a suitable subject for scientific study. E-language is roughly equal to Bloomfield's definition of language as "the totality of utterances that can be made in a speech community" (Bloomfield, 1926, p. 155) or to Lewis' (1975) equation of language to a pairing of sentences and meanings, used by a population holding certain beliefs and having an interest in communication. For structuralists like Bloomfield E-languages were the subject of linguistic study. But Chomsky alleges that E-language is an arbitrary, artificial construct, an epiphenomenon "that is understood *independently* of the properties of the mind/brain" (Chomsky, 1986a, p. 29, emphasis added) and that their study will "not reveal anything interesting about

the nature of language" (Ibid.). It is not clear to me how E-languages can be entirely independent of the properties of the mind/brain. Presumably, they have been generated by brains and are hence (on Chomsky's view) biological objects. But on the view advocated by Chomsky (1986a), language cannot be equated with the (potentially) infinite set of linguistic expressions that comprise E-language. He explicitly states that studying these expressions would not reveal anything interesting about the nature of language. What then *is* language?

Chomsky now defines I-language, the only object worthy of linguistic study, as "some element of the mind of the person who knows the language, acquired by the learner and used by the speaker-hearer" (Chomsky, 1986a, p. 22). We can see here that Chomsky has moved from the object of linguistic investigation in set theoretical terms to claiming that it is some well-defined part or module of the physical brain. Under this description the 'language faculty' is a module of the brain, and a person would literally have a language when this module of her brain was in a certain state S_s.

> The language faculty is a distinct system of mind/brain, with an initial state S_0 common to the species... Given appropriate experience, this faculty passes from state S_0 to some relatively stable steady state S_s, which then undergoes only peripheral modification (say acquiring new vocabulary items). The attained state incorporates an I-Language (it is the state of having or knowing a particular language). UG is the theory of S_0; particular grammars are theories of various I-languages (Chomsky, 1986a, p. 25).

Chomsky claims that the shift in focus from E- to I-language is a move towards realism because while E-languages "are not real-world objects but are artificial, somewhat arbitrary and perhaps not very interesting constructs" (Ibid., p.

26), I-languages understood as "the steady state of knowledge obtained ... are *real elements of particular minds/brains,* aspects of the physical world" (Ibid., p. 26, emphasis added). Chomsky calls this move a shift to a "mentalist interpretation of the study of language" (Ibid., p. 40), which allows for "the incorporation of the study of language within the natural sciences" (Ibid.). Chomsky later clarified that there is no difference between the language acquisition device and universal grammar:

> They are just two ways of looking at the same thing. Universal grammar is the name for the theory of the initial state of the language faculty. LAD is another name for the initial state, just looking at it from a different point of view. So *there's no difference*. (Chomsky, 2000b, p. 54, emphasis added)

When asked what the nature of the LAD is, Chomsky replied: "Well whatever the nature of language is, that is what it is" (Ibid.). If the nature of the LAD and the nature of language are the same it would follow that the LAD *is* language. This move seemed to address Postal's challenge. If we study language or UG or LAD or I-language, we study a biological object that is accessible to the natural sciences.

However, this reply is problematic. Chomsky has never abandoned the study of sentences and their grammatical properties. Chomsky commits himself (sometimes in the same publication, e.g. Chomsky, 1986a, 1995, 2000b, 2007b) to holding both; that language literally *is* a part of the physical brain and that language literally *is* an infinite set of linguistic expressions. As previously explained the inconsistency is not resolved by claiming the set of linguistic expressions is only potentially infinite as long as this set is literally located in the brain. To see why this is the case we have to look again at his writings. He writes: "the most elementary property of the language

faculty is the property of discrete infinity; you have six-word sentences, seven-word sentences…you can have ten-word sentences, twenty-word sentences and so on *indefinitely*" (Chomsky, 2000b, p. 51, emphasis added), and he explains that the language faculty *is* a physical part of the brain that has an initial state and can obtain different states depending on input (Ibid, p. 55). The property of having six- or seven-word sentences is a property that can apply to a set of sentences. This set can (potentially) be infinite. The property of having a well-defined initial state is a property that can apply to a (finite) physical organ. But Chomsky claims that the language faculty has both properties. Taken together the two claims are inconsistent.

A defender of Chomsky may reply that, while a literal reading of Chomsky's view is internally inconsistent, a more charitable interpretation resolves this problem. On this charitable reading Chomsky claims that language *is* a subsystem of the brain that *produces* a (potentially) infinite set of linguistic expressions. Clearly, this claim is not immediately internally inconsistent. Assuming that it is not, this move is not unproblematic. Chomsky repeatedly emphasized the similarity of the language faculty to other biological organs, such as the immune or digestive system (Chomsky, 1999) or the circulatory system (Chomsky, 2000a). Insisting on this analogy raises another problem for Chomsky's view. We know that the 'outputs' produced by these other organs are finite physical objects (enzymes, acids, blood cells, etc.). These products do not, even potentially, approach infinity. Thus, any talk about potentially infinite production in these biological organs can only be a metaphor. However, there are clear indications that, when Chomsky speaks about the language faculty, he does not speak metaphorically: "An I-language is a computational system that generates

infinitely many internal expressions, each of which can be regarded as an array of instructions to the interface systems, sensorimotor (SM) and conceptual-intentional (CI)" (Chomsky, 2007b, p. 3). Earlier in the same publication Chomsky has defined I-language as a state of the faculty of language (FL). FL is "a 'cognitive organ', … shared among humans and in crucial respects unique to them, hence a kind of species property" (Chomsky, 2007b, p. 1). This organ that is shared among humans must be a physical system, presumably a part of the biological brain. And, according to Chomsky this physical organ "generates infinitely many internal expressions". He does not indicate that the generation is potential but continues to claim that each of the infinitely many internal expressions "can be regarded as an array of instructions".

If the human language faculty generates infinitely many *internal* expressions, there must be a crucial difference between organs like the digestive or circulatory system on the one hand and the language faculty on the other hand. And it is precisely *this difference* that a meaningful theory of language needs to explain. Chomsky's emphasis of the *similarities* between the language faculty and other organs diverts attention from the fact that the fundamental difference between these organs is in need of an explanation.

To date Chomskyan science has done little to address this difference. He has re-iterated what we knew since Descartes; *that* language is unbounded and creative. He has repeated what Humboldt said; *that* language makes infinite use of finite means. But his work has not contributed any significantly new insights to answer the question: *how* does the biological language faculty produce a potentially infinite set of linguistic expressions? Chomsky claims that "the 'cognitive revolution' of the

1950s… [focused attention on the search for] the inner mechanisms that enter into thought and action…that generate expressions" (Chomsky, 2000a, p. 5). Yet, 60 years later we still know very little about the biological properties of these mechanisms. The areas of the brain that are implicated in language processing do not have any biological properties that are not found elsewhere in the brain. And recently Chomskyans have expressed pessimism about our ability to make significant progress: "If reason … has the biological basis it seems to have, reason must have its limits. Those limits are revealed in an incapacity to make scientific sense of the creative aspect of language" (McGilvray, 2009, p. 34). If this is true, then we are not closer to finding a scientific explanation for this creative aspect of language than Descartes was.

I will discuss next how Chomsky has attempted to use the E-/I-language distinction to explain language acquisition. Again, we will see that this distinction has made no substantial contribution to our understanding of language acquisition. In Chomsky's framework 'language learning' is really a misnomer because the child does not 'learn' anything. Rather, she is exposed to a finite array of linguistic data, and her "mind (incorporating S_0) constructs an I-Language that assigns a status to every expression" (Ibid., p. 31). This is possible because the language faculty, "a computational system that is rich and narrowly constrained in structure and rigid in its essential operations" (Ibid., p. 43), facilitates language acquisition. Chomsky alleges that an innate structure of this kind is necessary to solve 'Plato's problem'. This is, loosely speaking, the problem of explaining how children can 'know' so much more (about language) than they could have 'learned' from the input that was available to

163

them. Not only is the input insufficient to explain the output; it is also hugely variable (the actual sentences individual children hear vary greatly). Yet, the resulting language competence is more or less uniform, according to Chomsky: "irrespective of questions of maturation, order of presentation or selective availability the result of language acquisition is as if it were instantaneous ... intermediate states do not change the principles available for interpretation of data at later states" (Ibid., p. 54). Thus, while the data seem to play some role in language acquisition, what is 'doing the work' is the maturation of a genetically predetermined language faculty:

> our knowledge is richly articulated and shared with others from the same speech community, whereas the data available are much too impoverished to determine it by *any* general procedure of induction, generalization, analogy, association *or whatever*. There is good reason to believe that children learn language from positive evidence alone (corrections not being required or relevant), and they appear to know the facts without relevant experience in a wide array of complex cases ... *It must be*, then that the guessing instinct submits very few admissible hypotheses to the evaluation procedure. (Chomsky, 1986a, p. 55, emphasis added)

The language in this paragraph is clear: *all* potential methods of learning (any general procedure and whatever) are insufficient to explain language acquisition. Therefore, the 'guessing instinct' must restrict the admissible hypotheses. Here Chomsky does not only suggest a testable empirical hypothesis. He also creates the impression that the guessing instinct is necessary to account for language acquisition. However, he does not explain about the relationship between the positive evidence children receive (sentence/expression tokens) and the nature of the hypotheses that are evaluated by the evaluation procedure. And, as we shall see in section 3.5, even

the references to the biological entities in the brain that underwrite language are so vague, that no specific research program has ever been developed to search for these putative brain structures.

According to Chomsky (1995), an important upgrade to the LAD theorizing was the 'Principles and Parameter' (P&P) component. The earlier versions of Chomsky's theory were based on the assumption that a language consists of rules for forming grammatical constructions (such as relative clauses, passive, etc.). Chomsky explains that P&P broke with this tradition because

languages have no rules in anything like the familiar sense, and no theoretically significant grammatical constructions except as taxonomic artifacts. There are universal principles and a finite array of options as to how they apply (parameters), but no language-particular rules and no grammatical constructions of the traditional sort within or across languages (Chomsky, 1995, p. 5-6).

Chomsky claims that the cognitive system for individual languages consists of a computational system and the lexicon. The items of the lexicon belong to two categories that occur universally throughout the world's languages. The substantive category contains 'noun', 'verb', 'adjective', 'particle'; the functional category contains tense markers, complementizers etc. Parameters are restricted to formal features of language and possibly to formal features of functional categories. "In this context, language acquisition is ... the process of fixing parameters of the initial state in one of the permissible ways. A specific choice of parameter settings determines ... an I-language" (Ibid., p. 6). Chomsky proposes that this shift allows us to address the question of explanatory adequacy more rigorously than previous models because P&P

allows us to examine "how values are set by experience for finitely many universal parameters" (Ibid., p. 7).

However, the research throughout the years following the introduction of the P&P model has not produced any verifiable results regarding the process of parameter fixing. These failures have led to another 'far reaching departure' from former models (Chomsky, 1995, p. 10), the Minimalist Program. Chomsky proposes that the language faculty consists of a cognitive system and external performance systems. The cognitive system is comprised of a computational system and a lexicon for storing information. The performance systems are the articulatory-perceptual and conceptual-intentional system. These are interacting with the cognitive system at the phonetic form and logical form interface levels. According to Chomsky, "there is a single computational system C_{HL} for human language and only limited lexical variety" (Chomsky, 1995, p. 7). This computational system consists of two operations: Merge (the repeated combination of two syntactic objects) and Attract/Move (the movement of syntactic objects, p. 378). The Minimalist Program is a theory of Universal Grammar (UG) that considers a linguistic expression (SD) to be "the optimal realizations of the interface conditions, where 'optimality' is determined by the economy of UG" (Ibid., p. 171). The economy conditions of UG "select among convergent derivations" (Ibid., p. 378). This program does not assume the existence of any conditions relating lexical properties and interface levels (Ibid., p. 220), and Chomsky claims that the economy of UG is primarily concerned with derivational operations of the computational human language system (C_{HL}) like Merge and Attract/Move.

A very important departure from previous models is that in Minimalist Program "there are no levels of linguistic structure apart from the two interface levels PF and LF, specifically no levels of D-Structure or S-Structure" (Chomsky, 1995, p. 219). The elimination of D-Structure deserves special attention because it has been urged decades earlier by co-workers of Chomsky.

> …it has been a primary view of [Generative Semantics] writings that no real evidence has been attested supporting the existence of a level of "deep structure" distinct from logical structure, and, even more fundamentally, that there is no reason to expect any to be discovered (Postal, 1974a, p. 367).

It is common practice in science to acknowledge the earlier contributions of others, especially of former co-workers. Yet, we do not find any reference to this earlier work (e.g., Lakoff & Ross, 1976, McCawley, 1976) in *The Minimalist Program* and subsequent publications by Chomsky (for discussion see Levine & Postal, 2004). Like the previous 'updates', the minimalist program was hailed as *the solution* to problems that have plagued earlier versions of Chomskyan linguistics. Chomsky promised that minimalism would result in "a substantially different conception of the mechanisms of language" (Chomsky, 1995; p. 219). Specifically, it would move towards "a theory of language that takes linguistic expressions to be nothing other than the formal object that satisfies the interface conditions in the optimal way" (p. 171). Under this view linguistic expressions are not physical objects. But, given that Chomsky has not abandoned the view that I-languages are part of our biological endowment, it is again unclear how the brain contains and/or

produces these linguistic expressions. And it becomes increasingly difficult for researchers to interpret what exactly Chomsky's view amounts to. Even generative linguists expressed severe criticism of the newest addition to the Chomskyan theory-collection:

> The advent of minimalism in the mainstream of syntactic theorizing highlights an interesting shift in scientific values. At least from the Aspects theory through Principles and Parameters theory it has often been remarked that the syntax of natural language has some surprising, or at least abstract, non-obvious properties. One example is the transformational cycle, another is rule ordering, another is the ECP, and so on. Such properties are not predictable on the basis of 'common sense', and do not appear in any way to be logically necessary. The fact that they appear to be true of natural language thus tells us something, albeit indirectly, about the architecture of the language faculty in the human mind/brain… With the M[inimalist] P[rogram] we see a shift to a deep skepticism about formal devices of traditional syntactic theory that are not in some sense reducible to 'virtual conceptual necessity'. Such a perspective thus explicitly rules out precisely the major theoretical achievements of the past. *All of them*." (Culicover, 1999:137-138, emphasis added)

Here Culicover does not merely critique a detail of the new program but challenges the notion that Chomskyan theorizing has followed a coherent path. In fact, the critique goes even further and suggests that Minimalism entails that the theoretical achievements of Chomsky's career have been illusionary. Others have expressed similar views. Especially the contribution of the minimalist program to genuine progress has been questioned: [MP] did not generate new analytical tools, and thus failed to generate novel ways of looking at well known paradigms or to expand and solve old problems, an essential ingredient for progress made at this point" (Koopman, 2000, p. 2). Another linguist expresses his concerns in similar

terms: "The Minimalist Program … fails to satisfy basic scientific criteria such as respect for data, unambiguous formulation, falsifiability" (Seuren, 2004, p. 4). These concerns, which are representative for those expressed by others (e.g., Pullum, 1996; Newmeyer, 2003; Postal, 2004; Culicover & Jackendoff, 2005; Pinker & Jackendoff, 2005), do not concern details of the Minimalist Program but challenge its legitimacy as a scientific program.

Given some of Chomsky's claims, this seems to be a fair challenge. He announces, "The end result [of the Minimalist Program] is a substantially different conception of the mechanisms of language" (Chomsky, 1995, p. 219) but has to admit that "current formulation of such ideas still leaves substantial gaps and a range of plausible alternatives" (Ibid., p. 221). And anyone but Chomsky might get somewhat concerned by this evaluation of what has been achieved so far with the new program: "On minimalist assumptions, these seem to be the right kinds of moves, though doubtless not yet the right ones. Like earlier efforts guided by the same goals, they raise many questions and, if plausible, call for a good deal of rethinking of what has been assumed" (Ibid., p. 379). Thus, seemingly, the achievements of the past 40 years of linguistic theorizing have resulted in the realization that the previous conceptions of the mechanisms of language were incorrect, that the new research program leaves open many gaps and that the current moves are probably not the right ones. This sounds like a very sobering summary of actual achievements.

Defenders of Chomskyan linguistics observe that "[m]isunderstanding of [Chomsky's] view is widespread" (Smith, 1999, p. 3) but imply that the problems lies with those who critique what they fail to understand. Chomsky's work properly

understood has followed a coherent path, and "progress was made …with the introduction of the minimalist program" (McGilvray, 2009, p. 29). However, the above stated admissions by Chomsky himself indicate that Culicover's criticism is not based on a misunderstanding of Chomsky's work. In addition Chomsky's prognosis for an adequate explanation of the alleged domain specificity of the language faculty does not sound particularly optimistic:

> Suppose that [the minimalist] approach proves to be more or less correct. What could we then conclude about the specificity of the language faculty (modularity)? *Not much*. The language faculty might be unique among cognitive systems or even in the organic world, in that it satisfies minimalist assumptions. (Chomsky, 1995, p. 221, emphasis added)

The argument is not very informative: if minimalism is right, then the language faculty *might be* the (only) system to satisfy minimalist assumptions. Such an argument does not support more than Chomsky's sobering conclusion: we will know not much about the modularity of the language faculty *if* minimalism is correct. What, then, the reader may ask, is the advantage of the Minimalist Program? Chomsky offers one answer: "…the Minimalist Program, right or wrong, has a certain therapeutic value" (Chomsky, 1995, p. 233). This is presumably a value for any research program. Researchers not in need of therapy, however, seem right to insist, that a *linguistic* research program should also offer specific linguistic hypotheses that make testable predictions. Chomsky continues to insist that his linguistics needs to be treated as part of the natural sciences (e.g. Chomsky, 2005, 2007 b, 2010b). It certainly is desirable to view linguistics as an integral part of a multidisciplinary

research project that aims to provide a better understanding of the human mind. For this reason practitioners of fields that are relevant to linguistics (e.g., developmental psychologists, neuroscientists, physiologists, computational and experimental language researchers) need to have a clear understanding of the implications of Chomsky's view to be able to find confirming (or disconfirming) evidence for it.

Hopes for clarification arose again when another update to Chomsky's theorizing was provided in 2002. Hauser, Chomsky, and Fitch co-published a paper that introduced a new, evolutionarily motivated, framework for the human language faculty. Here the authors argued that human language requires two interacting components: the faculty of language in the broad sense (FLB) and the faculty of language in the narrow sense (FLN). FLB includes a sensorimotor system that is responsible for producing and receiving linguistic signals (speaking and hearing) and a conceptual-intentional system that allows the organism to categorize and organize information and to understand social cues. FLN, which is really a subsystem of FLB, "is the abstract linguistic computational system alone, independent of the other systems with which it interacts and interfaces. FLN is a component of FLB, and the mechanisms underlying it are some subset of those underlying FLB: "…The core property of FLN is recursion… it takes a finite set of elements and yields a potentially infinite array of discrete expressions" (Hauser, Chomsky, & Fitch, 2002, p. 1571).

This account addresses some questions regarding language evolution and has been acknowledged accordingly. Focusing on brain structures, the authors claim that several species have analogous systems to FLB but FLN is unique to humans. They consider "the possibility that FLN has evolved for reasons other than language" (p.

1573). On this view recursion could occur in animal navigation and number-quantification systems and could support some of the intra-species social interactions. It would be possible that recursion systems in non-human animals are highly domain specific while the human recursion system seems to be domain general which allows its use for solving a much wider range of problems. Finally, Fitch, Hauser and Chomsky (2005) argue, "the putative absence of obvious recursion in one of [the human] languages … does not affect the argument that recursion is part of the human language faculty [because] …our language faculty provides us with a toolkit for building languages, but not all languages use all the tools" (pp. 203-204), and they suggest that "the contents of 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" (Ibid., p. 181).

In response to criticisms of work on the Minimalist program Chomsky proposes a strong minimalist thesis (SMT):

> An I-language is a computational system that generates infinitely many internal expressions, each of which can be regarded as an array of instructions to the interface systems, sensorimotor (SM) and conceptual-intentional (CI). To the extent that third factor conditions function, the language will be efficiently designed to satisfy conditions imposed at the interface; one can imagine more radical theses, to which I will briefly return. We can regard an account of some linguistic phenomena as *principled* insofar as it derives them by efficient computation satisfying interface conditions. We can therefore formulate SMT as the thesis that all phenomena of language have a principled account in this sense, that language is a perfect solution to interface conditions, the conditions it must at least partially satisfy if it is to be usable at all. (Chomsky, 2007b, p. 3).

This SMT is supposed to approach the problem of determining the character of the language faculty bottom-up, asking how little can be attributed to UG while still accounting for all I-languages that exist. This approach should eventually merge with the more traditional top-down approach that accounts for how much needs to be attributed to UG to explain language acquisition. But in the remainder of the paper no specific details about how this work is to progress are given. Chomsky claims some progress has been made but also admits: "Just how far this line of inquiry can reach, of course one cannot know. As it proceeds, it approaches more closely the goal of identifying the principles of UG, the residue that constitutes FL" (Chomsky, 2007b, p. 18).

Judging by the description that Chomsky "tries to reinvent the field each time he sits down to write" (Fiengo, 2006, p. 471), it appears safe to assume that further updates and 'radical breaks' from earlier theorizing will follow. Therefore, at this point a comprehensive evaluation of Chomskyan linguistics seems premature (for critical evaluation see Postal, 2004; for praise McGilvray 2005). From the foregoing it has become clear, though, that future accounts need to address the longstanding criticisms regarding clarity and accessibility. As the following quotes show, Chomsky has repeatedly recognized that critiques often misunderstand his position: "Searle has also missed the central point…" (Chomsky, 1975a, p. 215). "Lakoff presents a very confused picture of the issues that have been under discussion" (Chomsky, 1973). "[Cooper] has, however, … seriously confused the issues" (Chomsky & Katz, 1975, p. 71). "McCawley has missed the point of my remarks" (Chomsky, 1980b, p. 47). "The post-1979 shift that Pullum perceives is imaginary… Pullum's conclusions are

based on serious misunderstanding throughout" (Chomsky, 1990, p. 147). "Godfrey-Smith cites only the last phrase quoted, misreading it as … exactly the opposite of what the passage unambiguously states" (Chomsky, 1995, p. 2). "Putnam and others … reject…'the innateness hypothesis' [which they attribute to me] …I have never defended it" (Chomsky, 2000a, p. 66). "…there is a huge literature arguing against the innateness of language; there's nothing defending the thesis" (Chomsky, 2000b, p. 50). "[In Boden's account] every reference to me … is fanciful, sometimes even bringing to mind Pauli's famous observation 'not even wrong'" (Chomsky, 2007, p. 4).

What is significant about these comments is that they indicate problems that go far beyond the usual misunderstandings of *what is entailed* by an author's view. According to Chomsky these scholars have not even understood the basic points of his view or they have mistakenly assumed he put forward a hypothesis when he has done no such thing. Given that so many scholars completely misunderstand Chomsky, it would appear to be desirable for Chomsky to express his position in a way that can be easily understood. He often claims to have done this. Unfortunately, reading his publications carefully reveals another picture. It is not uncommon that Chomsky claims to have always stated one position unambiguously. Yet, we can find him also stating the opposite of this position. For example on the one hand Chomsky describes empiricists as follows: "empiricism insists that the brain is a *tabula rasa*, empty, unstructured, uniform at least as far as cognitive structure is concerned" (Chomsky, 1977) and "empiricist currents, that would have us believe that the human mind is empty, a *tabula rasa*" (Chomsky, 1980, p. 270). On the other hand Chomsky

denies vehemently 'allegations' that he accused empiricists of 'blank-slatism'. He claims that he has "repeatedly, consistently, and clearly insisted that all rational approaches to the problem of learning, including 'associationism' and many others...attribute innate structure to the organism" (Chomsky, 1983, p. 310) and that "rationalists and empiricists alike attribute innate structure and principles to the mind" (Chomsky & Katz, 1975, p. 70). Intellectual honesty would require that Chomsky take ownership of the statements he has made in writing, retract those that are contrary to his view and clarify his position.

There are other examples of misleading statements in Chomsky's writings. In one publication Chomsky attempts to create the impression that his theorizing has always followed a coherent path. Claiming that even his earliest work was dedicated to the study of I-language, he writes: "In my 1965 book *Aspects of the Theory of Syntax*, for example, there is no entry for 'language' in the index, but many entries under 'grammar,' generally referring to I-language" (Chomsky, 1986a, p.29). When the reader checks the 1965 publication, she will notice that, technically speaking, Chomsky's claim is correct, because there is no entry labeled 'language' in the index. However, there is also no entry labeled 'grammar'. When she searches for entries containing either 'grammar' or 'language', she finds numerous entries for both (e.g., grammar, generative; grammar, particular; grammar, philosophical; grammar, universal; language learning; language and linguistic theory; language use; languages, artificial). Further, when she progresses to an entry that defines generative grammar she reads: "A fully adequate grammar must assign to each of an infinite range of sentences a structural description" (Chomsky, 1965b; p.5). But this is the notion of

generative grammar that Chomsky calls incorrect in the 1986 publication because it referred to E-language and "reinforced inappropriate analogies to the formal sciences" (Chomsky, 1986a, p. 29). Thus the reader is left wondering why Chomsky makes the misleading 'language' vs. 'grammar' claim and which notion of 'grammar' is correct.

A confusion of a different kind arises when Chomsky draws a distinction between language acquisition and language use that he also attributes to Descartes. He claims that for Descartes the "growth of an organism …including the acquisition of language and other cognitive (ethical etc.) systems is mechanical" (Chomsky 2010b, p. 104). Chomsky suggests that Descartes held that only the creative use of language could not be explained based on mechanical philosophy: "the major problem that Descartes and his followers raised about the apparent limitations of machines, …*had nothing to do* with the acquisition of language and other 'mental systems' but rather with their use, an entirely different matter" (Ibid., emphasis added).

Several problematic issues arise here. First, there is no indication that Descartes drew the Chomskyan distinction between language acquisition and language use. Descartes held that animals (like magpies) could be taught to imitate some phrases, but he never called this process language acquisition. On the most charitable reading Chomsky confuses here speech and language. But it is even dubious that Descartes would have held that a machine could acquire speech. Second, for Descartes language and 'other cognitive faculties' were not literally located in the physical brain. Thus, their acquisition could not have been explained based on the

same 'deterministic physical principles' that account for the growth of an organism. I have discussed Descartes' view in some detail in chapter 2 and suggested that it would be impossible for him to hold that a machine can *acquire* a complete language and other cognitive faculties that distinguish human beings from the brutes. Finally, Descartes' arguments in the Meditations and elsewhere, that he can imagine his mind being completely separated from his body, would be unintelligible, if language were located in the brain. In this case Descartes' soul could not have formulated '*Cogito ergo sum*'.

Apparently Chomsky does not have any suggestion how we could inquire scientifically into language use: "… the Cartesian question of creative [language] use…remains as much of a mystery now as it did centuries ago and may turn out to be one of those ultimate secrets that will ever remain in obscurity, impenetrable to human intelligence" (Chomsky, 2010a, p. 29). This pessimistic outlook is not restricted to our ability to formulate a scientific theory of performance. When asked specifically about the significant body of doctrine on different aspects of language and the established results in his field of research, he replied evasively:

> almost everything is subject to question, especially if you look at it from a minimalist perspective; about everything you look at the question is: why is it there?… We just have to see…You'll get new perspectives, things will be reinterpreted…I still don't think we understand [island conditions]. There's plenty of data that aren't understood…Such problems about…On the other hand, something will remain stable…But I don't feel that one can really predict much. It's a young field, changes are taking place fast, there are lots of unexplained things. I am sure there are new perspectives we haven't thought of yet… there are just too many mysteries. (Chomsky, 2002, pp. 151-153)

These sobering admissions after decades of enthusiastic success-reports and

the many internal inconsistencies contribute to the widespread confusion regarding Chomsky's position. In addition his frequent reformulations of hypotheses have left interpreters wondering which commitments still apply and which have been eliminated. This confusion is exacerbated by the fact that many of Chomsky's early works have been reprinted virtually unchanged several times. While one Chomskyan means to express admiration when he called Chomsky's work an "intellectual odyssey" (Smith, 2000, p. vi), I believe linguistic research should not resemble an ill prepared, dangerous, circuitous journey that eventually returns us to the point of origin. To demonstrate that this impression is mistaken, one task ahead would be to provide a clear statement of Chomsky's position that eliminates existing misunderstandings and confusions and to show evidence for genuine progress. The burden to provide such evidence is on Chomsky and his followers.

3. 5. Chomsky's Arguments for Innateness of Language

In this section I will discuss some of Chomsky's arguments for the innateness of language. I will focus on two main pillars of his conceptual framework: the existence of a domain-specific language acquisition device (LAD) in human brains and poverty of the stimulus arguments, which allegedly support the claim that human language is underwritten by some form of innate structure and/or knowledge. We have seen how problematic the definition of 'innate' is. I will stay in the following sections as close as possible to Chomsky's texts to avoid misinterpretation of what he means when he uses the term 'innate'.

According to Chomsky (1977, 1981, 1985, 1988, 1991, 2000a), humans and only humans have the innate capacity to acquire language. He postulates a species-specific language faculty (also called language acquisition device (LAD) or Universal Grammar) as a largely genetically determined part of our biological endowment and suggests that facts about language acquisition support his view.

Chomsky points out that many aspects of the linguistic competence of an adult speaker could not be explained based on the linguistic 'input data' available to a child when she acquires language. Therefore, these aspects could never have been learned but must have been available innately. Chomsky suggests: "all children share the same internal constraints which characterize narrowly the grammar they are going to construct" (Chomsky, 1977, p. 28). Further supporting evidence for the innateness of language supposedly comes from empirical research showing that language acquisition is quite fast and has characteristic stages whose order and duration seem largely independent of environmental factors. It is also remarkable that certain errors never occur. Chomsky's explanation is strikingly Platonic: "The speed and precision of vocabulary acquisition leaves no real alternative to the conclusion that the child somehow has the *concepts available before experience* with language and is basically learning labels for concepts that are *already a part of his or her conceptual apparatus*" (Chomsky, 1988, p. 24, emphasis added). Here we also see, that unlike his in earliest work, Chomsky embraces now a commitment to innate semantic content.

When we combine these claims, we have the 'poverty of the stimulus' argument. This argument can be stated as follows:

(1) If we observe the input and output of a cognitive system C, and if the input alone cannot account for the output, then system C needs to contain an innate structure that is sufficiently rich to account for the difference between input and output.

(2) The input (information that is available to the language learner) is insufficient to explain the observable output (language performance) (poverty of stimulus).

(3) System C contains an innate structure that accounts for the difference between input and output (*modus ponens* from 1 and 2).

This is a valid deduction, and all that remains to be done, is to show that the premises are true. The first premise seems intuitively plausible. If there is a difference between input and output, then an internal structure needs to account for the difference. The second premise, the claim that children receive insufficient input to account for their linguistic competence, deserves some further elaboration. To this I turn next.

3.5.1. The Poverty of Stimulus Argument

According to Michael Hymers (2005, p. 6) "the poverty-of-stimulus argument is the most compelling" argument in support of Chomsky's LAD. This view is shared by other philosophers of language (e.g. Martin, 1986, Stainton, 1996, Smith, 1999;

McGilvray, 1999, 2005; Matthews, 2005). Chomsky supplies some intuitive evidence to support his claim that children who first acquire language have to rely on 'insufficient input'. If we were to gather all the input data they have available (language they are exposed to, formal and informal teachings) and compare them to the output data (the language they eventually speak), then we would encounter a significant discrepancy between input and output. Simply put: the input cannot account for the output. Therefore, there must be an internal structure that can account (combined with the input) for the output. This appears plausible. Upon closer inspection, however, several points of contention emerge.

First, some of Chomsky's examples do not really support his claim that children know how to interpret certain structures "without training and relevant evidence" (Chomsky, 1985, p. 9). For example, he uses as illustration the two sentences

(6) John is too stubborn to talk to Bill

(7) John is too stubborn to talk to (Ibid.)

in support of his claim that the meaning of these constructions could not be learned by analogy from sentence pairs like:

(4) John ate an apple.

(5) John ate (Ibid).

And he asks the rhetorical question: "How does *every* child know, *unerringly*, to interpret the clause differently in the two cases? And why does no pedagogic grammar have to draw the learner's attention to such facts?" (Ibid, emphasis added). It is rather obvious that a learner would not be able to use the relationship between (4) and (5) to uncover the correct interpretation of (7) from (6). But it seems to be at least worth asking why she would attempt to draw an analogy between these two pairs in the first place or why she would *only* attempt to find an analogy between these (kinds of) pairs. Learning by analogy can be successful only if there is an analogy to be discovered. Chomsky has implied that the seemingly obvious analogy between the two examples given is the only analogy that could be discovered by the child. But he has not ruled out that other cues contained in the input could be exploited. In chapter 5 I will discuss some of these possibilities. Thus, Chomsky's 'discovery' that something else must explain the success of the young learner by itself is no threat to a data-driven theory of language acquisition. The interesting question is not whether or not superficial analogies between random pairs of input sentences alone can accommodate language learning. Instead we need to ask if there are *no* relevant analogies between examples that are encountered by a child. These analogies could be explicit or only implicit in the input.

Further, Chomsky has not provided any evidence for his claim that *every* child *unerringly* knows how to interpret these clauses. We will see in chapter 4 that significant differences exist among the productions of young children. Thus we cannot assume *a priori* that the above example supports the postulation of an innate domain specific LAD that allows for language production that varies little across

language communities. Empirical research conducted so far suggests that learning the correct interpretation of examples similar to those discussed here from information implicit in the input cannot be ruled out.

In addition we would need to account for language knowledge that the child has previously acquired (which by itself might in fact be insufficient to explain her success) and for the rich context in which sentences usually occurs. So she does not depend exclusively on the linguistic information contained in a given sentence to interpret an utterance. In chapter 4 we will see that children begin already at a very early age to integrate cues from different sources to extract relevant information from the language input. By the time they tackle the constructions Chomsky discusses here they have already several years of language-learning experience for which Chomsky fails to account.

Chomsky (1977) holds that the main evidence available for a language learner is positive evidence. According to Gold's (1967) formal proof, positive evidence from a finite set is insufficient to learn a natural language like English. Chomsky (1977) distinguishes three kinds of evidence that could be available to the learner:

- *positive evidence*: correct expressions in a particular language, for example word order, irregular verbs, etc.
- *direct negative evidence*: explicit correction of sentence structure or words employed in linguistic performance. An adult corrects his own or the child's linguistic mistakes.
- *indirect negative evidence*: a linguistic principle or rule hypothesized by a

child, is not instantiated in any sentences that she hears.

Chomsky holds that the actual learning set includes essentially positive evidence (correct utterances a child overhears) while direct negative evidence (correction by adults) and indirect negative evidence are negligible. In chapter 4 I will discuss recent empirical findings that challenge this claim. In chapter 5 I will discuss computer simulations that have shown that it is possible to learn natural languages like English from positive evidence alone if the 'end-state' is language performance comparable to that achieved by human speakers. For these reasons Gold's proof is now considered by many researchers as not relevant to natural language learning (MacWhinney, 2004; Chater & Vitanyi, 2007; Christiansen & Chater, 2009; Edelman & Waterfall, 2008).

I agree with Chomsky (1977, 1980a, 1985) that only empirical evidence (which is often open to conflicting interpretation) can decide whether or not a child receives sufficient negative evidence for language learning. But I want to suggest that Chomsky overlooks a different problem that arises here for his view. If it is correct that sentences like (11) and (12) demonstrate that only a LAD can account for the speaker's competence, then the same should be true generally for features that are difficult to explain of *all* natural languages. For example the genderization of nouns in German and other languages cannot be learned from drawing obvious analogies between random input pairs. Therefore, it would require precise parameter setting of the LAD. Given the tremendous number of nouns, it appears implausible that a genetic mechanism could account for such a highly specific output. Instead, it seems

sensible to suggest that children learn the proper gender assignment purely through positive evidence (they hear the correct combination), which coincides with indirect negative evidence (they rarely/never hear incorrect versions). Because children could choose among three possible genders per noun, errors should be very frequent. But the gender assignments are extremely robust and support the possibility of learning explicitly based on positive evidence. The robustness of correct gender assignment has been confirmed recently in computational modeling (Christiansen, p.c., 2011). It seems that Chomsky draws our attention to isolated instances that support his view while he neglects to explore examples that might challenge it.

Finally, Chomsky's view that the input is not sufficient to account for the output seems to pre-suppose that we have *complete* access to the data of input and output. But is this really the case? Or could Chomsky mistake a subset of the input for the entire input? Language learning does not occur under formalized experimental conditions. Rather, it is a 'messy process', and not all the factors that play a role in this process are well documented. In chapter 4 I will deal with empirical research that casts serious doubt on the assumption that the input is as impoverished as Chomsky makes it appear.

3.5.2. Chomsky's LAD

Chomsky has insisted throughout the years that the study of language needs to be conducted as scientific study, just as the study of any other 'bodily organ' (e.g., Chomsky, 1966, 1975a, 1985, 2002, 2005, 2006). For the success of any scientific

study it is important to have a clearly defined research project. One of the main goals of Chomsky's project has been to uncover the underlying structure of language, the brain state a person is in when she has mastered a language. In this section I will inquire how Chomsky's quest for one very particular piece in the puzzle, the LAD, has progressed. Has the research effort in the past 40 years resulted in the discovery of domain specific brains structures that underwrite linguistic competence? Or has the program accomplished a more modest goal: the theoretical clarification of what kind of structure we are looking for?

As we have seen, Chomsky uses the poverty of the stimulus argument to establish that an innate domain specific LAD is necessary to account for language acquisition. Given the importance of the LAD for Chomsky's arguments, we should expect to find a detailed description of such a structure in Chomsky's writing. But the search for clear descriptions is surprisingly difficult. A few quotations should suffice:

> It holds that beyond the peripheral processing mechanisms; there are innate ideas and principles of various kinds that determine the form of the acquired knowledge, in what may be, a rather restricted, and highly organized way. A condition for innate mechanisms to become activated is that appropriate stimulation be presented. (Chomsky, 1965a, p. 16)

Here we are told merely that peripheral processing mechanisms (whatever these may be) alone cannot account for language and that some internal mechanisms become activated when appropriately stimulated. Nothing is said about the structure of the internal mechanism or the nature of the 'appropriate stimuli'. In a similarly vague manner Chomsky (1968) proposes that "certain aspects of our knowledge and understanding are innate, part of our biological endowment, genetically determined,

on a par with the elements of our common nature that cause us to grow arms and legs rather than wings" (Ibid., p.17). It is undoubtedly true that, on some general level, everything that cannot be explained through environmental factors is part of our biological endowment. But what are the mechanisms that account for the innateness of 'certain aspects of our knowledge'? Chomsky does not say. In 1975 we are offered another definition of Universal Grammar, the most essential component of the LAD (so essential that at times LAD seems to reduce to UG):

> Let us define 'universal grammar' (UG) as the system of principles, conditions and rules that are elements or properties of all human languages not merely by accident but by [biological] necessity... Thus, UG can be taken as expressing 'the essence of human language". UG will be invariant among humans. UG will specify what language learning must achieve, if it takes place successfully...Each human language will conform to UG, though it will have other properties as well, accidental properties. If we were to construct a language violating UG, we would find that it could not be learned... (Chomsky, 1975a, p. 38)

Unfortunately, the amount of information contained in this passage stands in inverse relation to its verbosity. All we are really told here is that all human languages have certain properties and that a language without these properties could not be learned. To find this claim informative one would need to assume that a system that does not have the essential properties of language could still be learned. We can perhaps imagine that some extraterrestrial creatures have the ability to learn such a system. However, on Chomsky's view human language *is* a part of the human brain. Thus, the potential abilities of hypothetical aliens would not be relevant to Chomsky's view. In the following passage Chomsky seems to be more specific about the 'certain aspects':

> These principles determine the type of grammars that are available in principle. They are associated with an evaluation procedure which, given possible grammars, selects the best one. The evaluation procedure is also

part of the biological given. The acquisition of language thus is a process of selection of the best grammar compatible with the available data. If the principles can be made sufficiently restrictive there will also be a kind of 'discovery procedure'. (Chomsky, 1977, p. 41)

We are still left in the dark about *how* the 'evaluation procedure' can be a part of the 'biological given' (whatever that means). Are there unique structures within the brain, language-specific enzymes or other chemicals? And where in the brain are these structures located? It has been confirmed that Broca- and Wernicke- areas are involved in language production and comprehension, but these areas do not contain any unique structures that are not found anywhere else in the brain (or in the animal kingdom). So what is it that makes the language faculty unique? This question seems legitimate because Chomsky denies that 'general purpose learning mechanisms' can account for language. Again, we are given a general description of language learning and are left without hints about the details of the mechanisms involved:

One may think of [the language] faculty as a "language acquisition device", an innate component of the human mind that yields a particular language through interaction with presented experience, a device that converts experience into a system of knowledge attained: knowledge of one or another language. (Chomsky, 1985, p. 48)

It does not become clear what the phrase 'an innate component of the human mind' means or what kind of interactions convert experience into knowledge of a language. All we are told is a variation of the poverty of the stimulus argument: input cannot explain output, therefore some innate structure exists, and this structure converts input into output: "We have by now, fairly substantial evidence that one of the components of the mind/brain is a language faculty, dedicated to language and its

use" (Chomsky, 1993, p. 4). But Chomsky continues to evade genuine explanations in the by now familiar fashion when he writes:

> It seems that the child approaches the task of acquiring a language with a rich conceptual framework already in place and also with a rich system of assumptions about sound structure and the structure of complex utterances. These constitute the parts of our knowledge that come 'from the original hand of nature,' in Hume's phrase. They constitute one part of the human biological endowment, to be awakened by experience and to be sharpened and enriched in the course of the child's interactions with the human and material world. In these terms we can approach a solution of Plato's problem, along lines not entirely unlike Plato's own, though purged of the error of pre-existence. (Chomsky, 1991, p. 12)

However, it remains unclear how Chomsky purges 'the error of pre-existence' from Plato's view. His own view seems to support a pre-deterministic understanding: "Language is not really something the child does; it is something that happens to the child placed in an appropriate environment, much as the child's body grows and matures in a predetermined way when provided with appropriate nutrition and environmental stimulation" (Chomsky, 1988, p. 43). So while for Plato the mind recollects what it already knew, for Chomsky the body matures in a predetermined way. For Chomsky:

> …the child has a repertoire of concepts as part of its biological endowment and simply has to learn that a particular concept is realized in a particular way in the language. So the child has a concept, say *climb* in some abstract sense with all its weird properties and has to learn that it is pronounced 'climb'… (Chomsky, 2000a, p. 42)

This is eerily reminiscent of the slave boy scene in *Meno*. But it does not offer

189

any account for how these concepts came to be part of our biological endowment or where in the brain they are located. Landauer & Dumais, (2004) are similarly dissatisfied with Chomsky's explanations when they observe: "The fact that the mechanism is given by biology or that it exists as an autonomous mental or physical 'module' (if it does), tells us next to nothing about how the mind solves the basic inductive problem" (p. 63). And it seems not entirely implausible that Chomsky uses his repetitive but uninformative reference to biological endowment to silence possible critiques of his view, not to offer new insights in the problem of language acquisition. Eventually Chomsky does provide more details about the LAD:

> 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. Unless the switches are set one way or another, the system does not function. When they are set in one of the permissible ways, then the system functions in accordance with its nature, but differently, depending on how the switches are set. The fixed network is the system of principles of universal grammar; the switches are the parameters... When these switches are set, [a person] has command of a particular language and knows the facts of that language: that a particular expression has a particular meaning, and so on. Each permissible array of switch settings determines a particular language. (Chomsky, 2000 b, p. 69)

When we strip this passage of technical jargon, we still remain in the dark about what is going on when a child acquires language. Chomsky does not elaborate what 'permissible ways' are or at what point the system 'recognizes' a given language. One would also assume that the 'switch box' depends on more or less accurate inputs. But one of the motivations for Chomsky's theory was that the input is

often corrupted because parents do not speak grammatically 'perfectly' all the time. How a switch box distinguishes between input from different languages and input consisting of correct and incorrect utterances of one language needs elaboration. And it would appear that once the switches are set, the speaker has little flexibility to change the possible framework of her language. This is not a problem for language transmission from one generation to the next. But languages themselves have evolved over time, and syntax has changed considerably in the process (Bichakjian; 1988). When we assume that the LAD depends on input to set the parameters, then such a syntax evolution becomes difficult to explain. The following passage does not help to answer any of these questions:

> Like other biological systems, [the human language faculty] has a modular structure. We can at once distinguish two components: a cognitive system that stores information, and performance systems that access this information … the cognitive component…has an initial state, close to invariant across the species, and undergoes changes of state until it pretty much stabilizes, apparently pre-puberty. We may refer to a state of the language faculty as a language… an I-language, where *I* is to suggest *internal*, *individual*, and *intensional*, … We say that a theory of the language faculty attains explanatory adequacy to the extent that it gives the initial state. The term *explanation* is appropriate in the sense that the initial state can be considered abstractly as a procedure that determines an I-language, given data, and thus provides an explanation for the properties of the I-language and the expressions it generates, given the *boundary conditions* of experience. (Chomsky 2000a, p. 28, original emphasis)

Again we learn nothing about what actually *is* going on in the language faculty. It seems that Chomsky's theory provides at best a description, not an *explanation*, of the properties of the I-language. To sum it up: we have a difference between input and output that necessitates the existence of a language faculty. The

language faculty is such that it can account for the difference between input and output. "What we seem to find is that [whatever is innate to language] is specific. There are properties of the language faculty, which are not found elsewhere, not only in the human mind, but in other biological organisms as far as we know" (Chomsky, 2000a, p. 31). This statement is utterly uninformative; if the language faculty *is* *specific* to language, then we should expect it to have properties that are found *nowhere* else. What is missing is an account of what these properties are. Another recent formulation is not any more informative

> An elementary fact about the language faculty is that it is a system of discrete infinity. Any such system is based on a primitive operation that takes *n* objects already constructed, and constructs from them a new object: [I]n the simplest case, the set of these *n* objects. Call that operation Merge. (Chomsky, 2005, p. 11)

Returning to the two questions asked at the beginning of this section we have to admit that little progress has been made towards answering them. Chomsky's research has not resulted in the discovery of any domain specific brains structures that underwrite linguistic competence. And he has not provided significant theoretical clarifications of the kind of brain structure we are looking for. At times he admits this: "…. [we can assume that] something of this sort [P&P approach] is correct, but without trying to be very clear about the matter, since too little is understood to venture any strong hypotheses" Chomsky, 1995, p. 6).

Not surprisingly this failure has led some critics to conclude that the scientific project of confirming the LAD hypothesis has failed. Michael Tomasello has worked over decades on language acquisition issues (e.g., Tomasello 1992a, 2000, 2003,

2004, 2005) and was not able to find empirical evidence in support for a Chomskyan language acquisition device. In a recent publication he summarizes empirical cross-linguistic research that suggest:

> …parameters as theoretical constructs have not worked empirically, and the immediate prospects for progress do not seem promising. Parameters assume a range of formally specified lexical and functional categories, whose cross-linguistic status is problematic …, and whose linking to particular languages in the cases where they do apply seems to be pretty much impossible. (Tomasello, 2005, p. 188)

For Tomasello the continuous problems with formulating hypotheses or confirming those that have been formulated indicate that the Chomskyan approach has failed. He "it has been four decades now since *Aspects of a Theory of Syntax*, and there is less agreement than ever before about precisely what is and what is not in universal grammar" (Tomasello, 2005, p. 194) and suggests that the only way out of the current impasse is to rely on a fundamentally different assumption about the nature of language. It is up to Chomsky to demonstrate that Tomasello's interpretation of the work on universal grammar is incorrect. Whether or not he will succeed remains to be seen. I do not want to speculate about future theoretical developments at this point. Instead I will turn in the next chapter to empirical research that has been conducted in recent decades and evaluate whether or not any progress has been made in this area.

3.6. Conclusions

In this chapter I have attempted to provide a coherent account of Chomsky's linguistic theorizing. This task has proved to be incredibly difficult. Especially, I have found little evidence for the allegedly steady progress of Chomsky's work towards theories of greater simplicity and clarity (McGilvray, 2009). Rather, it appears that his critics (e.g., Katz, 1981; Botha, 1989; Sampson, 1991; Seuren, 1998; Boden, 2006) are correct to point out that the clearest and most unproblematic expression of his position can be found in his early work (*Syntactic Structures*, 1957). Here he lays a foundation for a scientific research program for linguistics. Naturally, many of the details of this program could only be sketched at the time, but it appeared to be a solid promissory note towards future research. Yet, 50 years later many of the promises made in the 1950s remain unfulfilled. In this chapter I have identified several specific issues that require clarification.

First, Chomsky's work has not contributed to the clarification of the concept of innateness. While he has repeatedly cautioned that he has never defended an innateness hypothesis (e.g., Chomsky, 1986a, 1995, 2000a,b, 2002, 2005), he has never indicated that the concept of 'innateness' requires clarification. References to 'innate endowment' and 'innate resources' are prevalent in his writings (e.g., Chomsky, 1966, 1975c, 1980, 1986a, 1995, 2000a, 2002, 2007b, 2009a, 2010b), but Chomsky does not define what 'innate' means. To make genuine progress on the innateness issue it is essential to clarify what 'innate' means and which aspects of language are innate.

Second, after several decades of theorizing, the nature of Universal Grammar

and the language acquisition device remains vague. To accommodate the necessary collaboration with neuroscientists and brain physiologists Chomsky needs to specify the properties of UG and LAD. Pace Postal (2009) it is not necessary that Chomsky himself conducts experimental biological work. But he needs to provide clear conceptual guidance to those who do.

Third, the poverty of the stimulus argument needs to be stated in an unambiguous way. It needs to be clear in which way the stimulus is impoverished and why this particular limitation provides an insurmountable obstacle to domain-general language acquisition. To evaluate the poverty or richness of the stimulus it is essential to incorporate the results from recently completed work by computational and experimental language researchers. A precise formulation of the poverty of the stimulus argument could provide valuable guidance for the work of these researchers.

Fourth, in order to evaluate whether opponents of Chomsky's view are dogmatists (e.g., Chomsky, 1966, 1980, 1986a, 2002, 2005, 2007) or have addressed genuine limitations with Chomsky's account, it is essential to eliminate inconsistencies in the account. Here it will be especially important to evaluate what the language learner accomplishes, how long it takes her to acquire language, whether language competence is uniform across the species and whether it is impossible to acquire language from the input that is actually available.

Fifth, Chomsky needs to explain how finite physical brains can produce infinite outputs. Which are the properties that allow the language faculty to produce outputs that are potentially infinite? Chomsky needs to show how his account of language overcomes the inconsistent ontology Katz (1985, 1996), Katz & Postal

(1991), and Postal (2003a, 2004, 2009) attribute to it. Given the commitments of Chomsky,

> [Chomskyan] linguistics requires that there be infinitely many real objects to serve as referents for the linguist's statements about infinitely many sentences of a natural language, while a concrete linguistic reality guarantees that there are none for the infinitely many possible but never-to-be-actual sentences. (Katz, 1996, p. 279)

What is needed is a clarification of what 'language' is. If language is a biological organ (as claimed in Chomsky 1986a, 1999, 2002, 2005, 2007a,b, 2009a, 2010b) then it is finite. If language is a collection of potentially infinitely many sentences or expressions (as claimed in Chomsky 1957, 1965a, 1968, 1977), then finite human brains can at best instantiate a part of language. And if language is an abstract object (as claimed in Chomsky 1957; 1961, 1977) then the nature of the relationship between language and brains needs to be explained. Whether or not these points can be addressed remains to be seen.

Fortunately, Chomsky himself agrees that it is the responsibility of the proponent of a view to explain it in accessible terms. If that cannot be done, we are under no obligation to pay attention to it. He writes that, if he comes across a subject in science he's not familiar with but wants to learn about, he knows just what to do:

> I ask someone who concentrates in these areas to explain to me what I want to know at a level corresponding to my interest and understanding. And quite commonly they can do so to my satisfaction, at least if they have a serious grasp on their own field ... But when... the practitioners of these [fields] do not seem to be able to explain to me what is being said, or what insights they have reached... I see absolutely no basis for the claim that by failing to understand work that no one can make

comprehensible to me I am being 'insensitive' or showing lack of interest. (Chomsky, 2010b, p. 108)

Chapter 4. Language Acquisition

4.1. Introduction

In the previous chapter I discussed some of Chomsky's arguments for the necessity of an innate domain-specific LAD. The poverty of the stimulus argument was the main argument that he put forward in support of this conclusion. Chomsky holds that many aspects of the linguistic competence of an adult speaker could not be explained based on the primary linguistic input available to a child when she acquires language. Therefore, he concludes, these aspects could never have been learned but must have been available innately. Roughly speaking, the Poverty of the Stimulus Argument claims that the postulation of a LAD is necessary because general-purpose learning abilities plus stimuli available to infants are insufficient to account for the accuracy and speed of language learning and the complexity of the language learned (e.g., Chomsky, 1977; 1988; 2005; 2007; 2009). Chomsky has repeatedly responded to philosophical criticisms of this argument by stressing that he did not put forward a deductive argument that could be refuted by *a priori* reasoning but an argument to the best explanation that is supported by rich empirical evidence (e.g., Chomsky 1975a, 1985, 1993, 2000a, 2002).

Empirical evidence will be the focus of the following inquiry. Seemingly for Chomsky, empirical evidence that relates to the language faculty can be quite diverse: "Judgments about the interpretation of sentences (essentially, perceptual judgments) are data, along with the results of priming studies and electrical activity

of the brain. We seek any data that may provide evidence bearing on the nature of the language faculty" (Chomsky, 1994, p. 23). Over the past decades scientists have accumulated an impressive amount of evidence fitting Chomsky's criteria. In the following sections I show that the empirical evidence accumulated by scientists in various domains (e.g., developmental psychology, experimental and computational language acquisition research) is compatible with a data-driven general purpose model of language acquisition and does not support the conclusion that an innate domain specific LAD is necessary to account for language acquisition.

First, I will examine what is known about the power of general learning processes and about the incremental nature of the initial process of language learning. I suggest that we need detailed knowledge of all the steps that are involved in language acquisition before we can rule out the possibility that general-purpose learning mechanisms are sufficient to account for language acquisition. In addition I will examine the role of negative evidence for language acquisition. I will argue that the alleged a lack of negative evidence for an innate domain specific LAD is hardly conclusive evidence for it. Finally, I discuss studies that appear to show that early language input contains abundant statistical information. I will inquire whether this information could be sufficient for data-driven language learning.

4.2. Arguments for the Innateness of Language

As discussed in chapter 3, Chomsky (1968 - 2005) postulates a species-specific language faculty as a largely genetically determined part of our biological

endowment and claims that facts about language acquisition support this view. He holds that many aspects of the linguistic competence of an adult speaker cannot be explained by data-driven general-purpose learning mechanisms. Only "an innate component of the human mind that yields a particular language through interaction with presented experience" (Chomsky, 1985, p.5) can account for language learning. Chomsky asserts, "all children share the same internal constraints which characterize narrowly the grammar they are going to construct" (Chomsky, 1977, p. 28). He interprets empirical research, which shows that language acquisition is quite fast and has characteristic stages whose order and duration seem largely independent from environmental factors, as further evidence supporting the innateness of language. Chomsky holds that the actual learning set available to a child includes essentially only positive evidence (correct utterances a child overhears) while negative evidence (correction by adults) is negligible, and he concludes that an LAD is needed to account for the success of the young language learner. His explanation is strikingly Platonic: "The speed and precision of vocabulary acquisition leaves no real alternative to the conclusion that the child somehow has the *concepts available before experience* with language and is basically learning labels for concepts that are *already a part of his or her conceptual apparatus*" (Chomsky, 1988, p. 24, emphasis added).

At the heart of Chomsky's claims about language acquisition is what has become known as the Poverty of the Stimulus Argument. From the premises (i) that the input received by the language-learning child is insufficient to explain her output depending on data driven general purpose learning mechanisms alone and (ii) that, nevertheless, children acquire language competence with seeming ease, he draws the

conclusion that a domain specific innate structure (the LAD) must exist. Allegedly, this LAD allows children to infer rules of grammar from an insufficient inductive database and become linguistically competent speakers. Unfortunately, Chomsky remains vague about the specifics of the LAD (see chapter 3), and his terminology is not always consistent; LAD, generative grammar, and language faculty all seem to be semantically interchangeable. What appears to be subscribed to under the different 'LAD' definitions is that the 'LAD' is a genetically determined independent neurocognitive entity (Marcus, 2006) that is specific to our species and to the domain of language and that differs from general purpose learning mechanisms. This will be then the sense in which I use LAD here and the evidence presented here will relate to LAD understood in this way. I will not speculate about what kind of knowledge would permit the totality of the child's potential output, but I will concentrate on specific examples of input that *is* available to *beginning* language learners.

The claim that children receive insufficient input to explain their language-competence seems to be supported by ubiquitous anecdotal evidence and has been highly influential for many philosophers, linguists and cognitive psychologists. Lenneberg (1967), Martin (1987), Ramsey & Stich (1991), Stainton (1996), McGilvray (1999), Smith (1999), Musolino et al. (2000), Crain, & Pietroski (2001), Crain & Pietroski (2002), Legate & Yang (2002), Lidz et al. (2003), Lidz & Waxman (2004), Baker (2005), Lightfoot (2005), and Cattell (2006) agree that the Poverty of the Stimulus Argument lends strong support to a domain specific LAD. Like Chomsky, they hold that the input available to the language learner is indeed insufficient to explain the resulting language competence. In the next sections I

present empirical evidence that casts doubt on this widely held assumption.

4.3. Empirical evidence gathered during language acquisition

To date extensive domains of language acquisition and input data remain unstudied, despite the large quantity of research dedicated to gathering information about the input (Behrens, 2006). This is not surprising given the vast amounts of input. Cameron-Faulkner et al. (2003) estimate that an average toddler hears between 5000 and 7000 utterances per day (or 5.5 to 7.6 million from their first to their fourth birthday). Quantitative studies that record the complete language productions of an individual child over even very short time periods are rare (see e.g., Wagner, 1985).

It is even more difficult to collect complete input data. Van den Weijer (1999) recorded and analyzed most of the spoken speech[3] to which an infant was exposed between the ages of 6 and 9 months. Behrens (2006) recorded the language development of one boy over a three-year period (see also MacWhinney, 1995; MacWhinney & Snow, 1985, 1990; Sampson, 2002 for more extensive database collections). And yet it is only very recently that a researcher has begun to attempt to gather an uninterrupted record of all language input of even just *one* individual child (Roy, 2007). Clearly, we need much more work in this direction in order to understand what sorts of utterances constitute the typical input to children (Pullum & Scholz, 2002).

As the foregoing suggests at this time arguments based on categorical claims

regarding the language input are not supported by empirical evidence. Yet, many of the claims put forward in support of Poverty of the Stimulus arguments lend support to what Dawkins (1986) has called 'arguments from incredulity'. Far from establishing empirical facts they rely on 'unquestioned axioms' that the impoverished language input and the insufficiency of data-driven general-purpose learning mechanisms necessitate an LAD (Sampson, 2002). To date there are no actual measurements of children's performance limitations (Tomasello, 2000); instead it is assumed that "there is no way that a child could get from concrete item based linguistic structures to the powerful abstractions that constitute adult linguistic competence" (ibid, p. 246).

To give a realistic account of language acquisition we need to pay attention to the all the steps involved in early language acquisition (Tomasello, 2000). During infancy the child learns crucial information about language that will help her in the acquisition of more complex structures later on. Several researchers have suggested that the same mechanisms that allow infants to segment the continuous speech stream into individual words and to access meaning can also account for the acquisition of complex syntactic structures (e.g., Deacon, 1997; Brent, 1999; Pacton et al., 2001; Christiansen & Ellefson, 2002; Maye et al., 2002; Saffran, 2003).

It will become evident that (i) many of the cognitive processes involved in early language acquisition could be based on general-purpose learning mechanisms and (ii) the complex interactions between child and mother during language learning/teaching (Newport et al., 1977; Moerk, 1992; Tomasello, 2003; Cameron-Faulkner et al., 2003; Dale & Spivey, 2006) provide more than verbal support for

the young language learner. It seems to be a mistake to reduce the primary linguistic input to spoken language directed explicitly at the child. Children access a wide variety of linguistic input, and they make use of non-linguistic information when they learn language. Parents, caregivers, and peers provide verbal feedback (Moerk, 1992; Demetras et al., 1986; Gentner & Goldin-Meadow, 2003; Gelman et al., 2005) and non-verbal support such as pointing and eye gaze (Goldin-Meadow & Mylander, 1990; Tomasello, 1992; Deacon, 1997; Tomasello, 2003; McNeill, 2005). I will suggest that for an adequate evaluation of the available input all sources of input need to be accounted for.

Second, although it is true that children acquire language earlier and seemingly with greater ease than they do knowledge in any other abstract domain, they need to learn a considerable amount of information about the nature and organization of their native language long before they can begin to produce recognizable words (Jusczyk et al., 1999). The evidence that I provide in section 4.3.1.1 will show that language learning appears to be rather slow in the first 18 months of life if we measure it solely by vocabulary acquisition.

Third, in order to evaluate the poverty (or richness) of the input, all aspects of the input need to be taken into consideration. It has become evident that statistical regularities of language provide a wealth of implicit information to the young language learner. In section 4.4.1.2. I explore research showing that young infants can track statistical regularities of language (Saffran et al., 1996, Brent & Cartwright, 1996; Mattys & Jusczyk, 2001; Saffran, 2001, 2002, 2003; Maye et al., 2003). Studies with older children (e.g., Hall & Graham, 1999; Pacton et al., 2001; Diesendruck et

al., 2006) have shown that children continue to rely on statistical features of the language input. I will suggest that proponents of the Poverty of the Stimulus argument need to rule out that mechanisms such as statistical learning can account for the acquisition of human language. Furthermore, if infants can use statistical information in early language-learning tasks, then the language input may not be as impoverished as it appears at first glance. Only a careful analysis of all available input data can show whether or not the gap between input and output is insurmountable.

4.3.1. Language Acquisition During Infancy

In this section I discuss studies showing that there is considerable language input and language learning during the prenatal and early infancy periods. Further, it will become evident that it takes the infant several months to acquire abilities that are often considered to be innately available. This suggests that we need to re-evaluate how many of seemingly innate abilities are indeed inborn.

4. 3.1. 1. Early Beginnings

Historically, little attention has been paid to what very young infants may be able to learn because it was assumed that no language learning would occur this early in development (Reali & Christiansen, 2005). This assumption has led to the belief

that some of the early language-related abilities must be innate. It has been suggested that the finding that "infants a few days old who have been given minimal exposure to a particular language… already know that language's sound and have already begun to develop one natural language rather than another" (McGilvray, 1999, p. 66) provides empirical evidence supporting the existence of an innate LAD. McGilvray argues that these young infants could not have acquired the information they possess from their environment because their exposure to (spoken) language has been negligible up to this point. Similar arguments have been put forward by Smith (1999), Petitto (2005), Cattel (2005), Matthews (2005), Isac & Reiss (2008), Hornstein & Boeckx (2009), and Chomsky, (2002, 2009a, 2010b).

However, empirical data concerning the earliest steps of language acquisition lend support to an alternative hypothesis. It is now believed that language learning begins very early in life. According to this view, the recognition of some features of the native language at birth can be explained by pre-natal exposure to this language. Infants begin to receive rich information about their native language through exposure to spoken words when they are still *in utero*. Most fetuses begin to respond to sound at 22 to 24 weeks (Hepper & Shahidullah, 1994), and by the time babies are born their basic auditory capabilities are relatively mature (Lasky & Williams, 2005; Saffran et al., 2006). At the normal frequency of the human voice (125-250 Hz) there is little attenuation by the mother's skin and tissues, and the fetus can hear the mother talking (for reviews see Hepper, 2002; Lasky & Williams, 2005) and is affected by exposure to other external sounds (Lecanuet & Schaal, 1996).

By the time infants are born they are able to discriminate the voice of their

mother from those of other women (DeCasper & Fifer, 1980). In addition, newborns also seem to be familiar with rhythmic properties of their native language. DeCasper & Spence (1986) further demonstrated that newborns showed a preference for passages of prose that their mothers had read aloud during the last six weeks of pregnancy. This preference persisted when the passages were read by another person to the newborn, suggesting that the child recognized not only the voice of the mother (as in DeCasper & Fifer, 1980), but also other acoustic properties of the prose. It seems unlikely that a genetically determined mechanism could account for this very specific preference. Instead, it is plausible to assume that the pre-natal exposure explains the post-natal preference. Thus, seemingly innate preferences can be explained on the basis of learning from pre-natal input (see also Moon & Fifer, 2000).

It is now believed that newborns can distinguish between utterances from languages that differ in rhythmic structure based on prenatal exposure to spoken language. Nazzi et al. (1998) showed that French newborns can discriminate between two unknown languages from different rhythmic classes (English versus Japanese), but they cannot discriminate between languages from the same rhythmic class (English versus Dutch). Over the course of several months infants learn to discriminate their own language from other languages in the same rhythmic class (Nazzi et al., 2000a).

Between 6 and 12 months of age infants fine-tune the perception of the individual sounds that distinguish between words (or phonemes) in the language to which they are exposed. Werker and Tees (1984) found that 6- to 8-month-old babies distinguish between a wide range of sound differences that signal changes in meaning

either in their native language or in non-native languages. And, while some of the 8- to 10-month-old infants were still able to discriminate non-native language contrasts, virtually all 10- to 12-month-old infants discriminated only native language contrasts. It has been suggested that these changes in perception reflect the growing ability of the infant to focus her attention only on those acoustic dimensions that are relevant for her native language (Maye et al., 2002). I want to stress that it takes almost one year for the infant to learn to attend to several acoustic cues that will help her to distinguish individual words. This means that the infant acquires step-by-step the ability to make perceptual discriminations that she did *not* have innately. It remains to be determined whether the learning of native language contrasts is based on data-driven general-purpose (perception) or language specific (LAD) mechanisms.

On her journey towards language, the infant needs to learn what words mean and how they can be combined into utterances. The first steps towards these goals are taken early. Researchers have shown that young infants can use acoustic and phonological characteristics to distinguish between the two most fundamental syntactic categories in human languages: function words that carry information about grammatical relationships between words in a sentence (such as prepositions, postpositions, auxiliaries, and determiners) and content words, which have more concrete lexical meanings (such as nouns, verbs, adjectives, and adverbs). Across the world's languages, content words are more salient both acoustically and phonologically. They tend to be longer, to have fuller vowels, and to have more complex syllable structure than function words (Shi et al., 1998). Shi et al. (1999) habituated newborns to either content or function words. The newborns preferred to

listen to new words in the category that they had not heard before. Further, the results persisted regardless of whether the infants had heard English prenatally or not. These findings suggest that perceptual sensitivities allow infants to detect the acoustic and phonological characteristics that distinguish the two broad syntactic categories. It seems that we have here a case in which a general-purpose ability (acoustic perception) can be recruited to support a domain-specific task that might be taken to reflect advanced grammatical knowledge (word-category discrimination[8]). This offers an alternative to a domain-specific LAD.

Similarly, perceptual abilities might also support later word learning. As infants get older they seem to develop a preferential bias for listening to content words. Shi and Werker (2003) demonstrated that 6-month-old infants who were growing up in an English or Chinese environment prefer to listen to English content words over function words. Because the preference for content words was evident in children who were not exposed to the language of testing, Shi and Werker argue that this preference is based on the acoustic and phonological salience of content words. The researchers propose that the emerging preference for content words could lead to a listening bias that allows infants to learn more about content words, and to use this knowledge in subsequent language acquisition tasks.

4. 3.1.2. Emerging Language Production in Young Children

Children need to master several cognitive skills in order to acquire language. One of these skills is the ability to produce the sounds of their native language and to

combine them into words and eventually into grammatically correct sentences. In the following section I touch on some of the abilities that the young language learner acquires before she produces her first meaningful sentences.

The production of words is preceded by a phase of vocalization during which the infant identifies, acquires, and practices the sounds that are common in her language. Babbling occurs at 6- to 10-months of age and prepares the child for producing the relevant set of sounds of her native language (for a review see Werker and Tees, 1999). The babbling child initially produces a wide range of sounds and later narrows this range to the sounds of her own language. This language-specific narrowing was shown by Boysson-Bardies and Vihman (1991). They found that the productions of 10-month-old infants exposed to one of four languages (French, English, Cantonese, and Swedish) were acoustically significantly different and that adults could reliably determine which productions were from languages other than their own.

One might ask how this honing takes place, particularly given that the parents commonly do not correct their babbling child (she receives virtually no negative feedback). Some researchers have shown that parental feedback can modify the phonological features of babbling (Goldstein & West, 1999; Goldstein et al., 2003), but we are not aware of any evidence that parents explicitly *correct* their babbling infant. Without receiving "explicit or focused training" (Vallabha et al., 2007) the child succeeds in refining her perception so as to categorize sounds along dimensions relevant to her native language.

Around the first birthday, most infants speak their first meaningful words, and

they only gradually expand their productive vocabulary. Initially children go through a single-word stage. They use a single word to express a complete thought to the listener: For example, when the child says *Milk*, she could mean a number of things: "*I want milk, I spilled milk, Mom has milk,* or *That's milk*…the holophrase derives its meaning in part from the context in which it occurs" (Jay, 2002, p. 367f). In the earliest stages of language production child meaning can differ greatly from adult meaning. Young children seem to focus more on context than on individual words. This may assist them in recognizing salient features in their environment and thereby make it easier to map objects onto individual words.

Proponents of the Poverty of the Stimulus Argument often downplay the role of parents or caregivers in language development. Instead, they hold that many aspects of language are known without training or explicit instruction (Chomsky, 1968, 1975a, 1985; McGilvray, 1999; Smith, 1999; Crain, & Pietroski, 2001; Pietroski & Crain, 2005). This claim has some intuitive appeal. However, before we draw the intended conclusion (that language acquisition must be supported by an innate LAD) we need to remember that children are exposed to a wealth of linguistic input and that parents adapt to the changing needs of the young language learner (Moerk, 1992). For example, many parents use child-directed speech to assist the child learning the meaning of early words: they emphasize individual words, reduce grammatical complexity and refer mainly to objects in the infant's environment (Moerk, 1992; Valian, 1999; Tomasello, 2003). Some recent research indicates that, compared to adult-directed speech, the information in child-directed speech seems to increase the speed with which infants acquire their vocabulary (McMurray, 2007).

Child-directed speech differs greatly from adult-directed speech. Van den Weijer (1999) documented that the lexical diversity of child-directed speech was less than 50% of that of speech directed to adults or older children. Cameron-Faulkner et al. (2003) found that parents repeatedly use utterances that include short sentence frames when addressing their infants. Fernald and Huerto (2006) demonstrated that 18-month-old infants are able to make use of these short, familiar sentence frames in word recognition and Mintz (2006) showed that a learner who is sensitive to frequent frames might be able to use this sensitivity to discover more complex syntactic structures. My point is that not everything that helps the child to learn language needs to have the structure of formal teaching.

In addition, children do not learn only from speech that is directed at them but also from language input that they overhear (Akhtar et al., 2001; Scholz & Pullum, 2002), such as conversations between adults or other children. This finding helps to explain how children can learn even in the absence of child-directed speech (as occurs in some cultures, see e.g., Ochs, 1985 as cited in Lieven, 1994). And yet, even the absence of child-directed speech does not require the invocation of an LAD. As Lieven (1994), Akhtar et al. (2001) and others have noted, in addition to overhearing language, many of the other cues that we discuss here, such as frequency and co-occurrence, could support language learning. Again, it is important that we account for the entire language input before we can decide whether or not an LAD is needed to explain language acquisition.

Empirical research has shown that the initial pace of vocabulary learning (from birth to 18 months) is modest. It has been suggested that during this time

children acquire and practice many cognitive abilities. An infant needs to be able to *see* object boundaries before she can form the hypothesis that ostensive definitions apply to whole objects. She needs to be able to *perceive* similarities and differences between objects before she can categorize them. Further, she needs to be able to resolve the conflict between the mutual exclusivity assumption (one name for one object, e.g. 'dog' for the family pet) and the need for taxonomic categorization (e.g. 'dog' for any dog-like object). Children acquire and practice these abilities over an extended time period. They learn to categorize the world and to understand how words refer to objects, actions, and properties. One hypothesis suggests that, once the child has acquired this knowledge, she can slot with ease new words into existing categories (Deacon, 1997). According to this view, general learning mechanisms could account for language acquisition, and an LAD would not be needed.

The fast acquisition of vocabulary (vocabulary spurt) and syntax after the second birthday is frequently used as supporting evidence for the existence of language-specific learning mechanisms that mature at genetically predetermined times (e.g., Chomsky, 1975a, 1985; Lightfoot, 1989; Pinker, 1994; Smith, 1999). However, recent research suggests an alternative account. McMurray (2007) conducted a series of simple computational simulations of vocabulary learning. Based on the assumptions that (i) words are learned in parallel and (ii) words vary in difficulty, he developed a model that replicated the vocabulary spurt. On his view the vocabulary spurt is an inevitable result of the infant's immersion in words of varying difficulty. A similar model was developed by Dale and Goodman (2005), who reported that the rate of change in words acquired is a function of the number of

words already learned. Finally, Ganger and Brent (2004) applied a detailed, quantitative method for detecting a vocabulary spurt in individual children (identifying an inflection point in the learning-rate curve) and found that only 1 out of 5 children goes through a vocabulary spurt. Findings like these suggest that we must rethink whether the vocabulary spurt provides supporting evidence for the existence of an innate language faculty that is shared by all members of the human species.

As children expand their vocabulary, they begin to construct syntactically complex multiword utterances. Again, some researchers have suggested that genuine syntactic progress and creativity in language use continue to be rather modest even after the child has acquired a sizable vocabulary. Lieven et al. (2003) analyzed the multi-word utterances produced by a 2-year, 1-month-old girl interacting with her mother. Of the 295 multi-word utterances recorded, only 37% were 'novel' (they had not been produced in their entirety before). A total of 74% of the novel utterances differed by only one word from previous utterances. 24% of the novel utterances differed in two words, and only few of the remaining utterances were more difficult to match. This suggests that the creativity in early language "could be at least partially based upon entrenched schemas and a small number of simple operations to modify them" (Lieven et al., 2003, p. 333). Similar results have been reported by Tomasello (1992b), Rubino & Pine (1998) and Tomasello et al. (1997). Such findings imply that the child devotes extensive time to practicing and rehearsing familiar utterances.

At around 18- to 24-months of age children begin to combine words into coherent utterances and the first two-word combinations appear. While it is difficult

to uncover how children learn the main grammatical categories of verb, noun, adverb, and adjective, there are some indications as to how they do so. As we noted earlier, there is some evidence that content and function words can be distinguished by their phonological properties (Morgan et al., 1987; Cutler, 1993; for a recent review see Monahagan et al., 2005). Even within the content category nouns tend to have more syllables than verbs in English, and, for bisyllabic words, first-syllable stress is generally a property of nouns, whereas verbs tend to have second-syllable stress (Kelly, 1992; McMurray & Aslin, 2005). Moerk & Moerk (1979) suggested that young children could learn some grammatical forms (e.g., past tense of verbs) by imitation even before they fully understand the function of these forms. This might allow the child later to recognize the function based on acoustic properties of the word. In the next section we will see that children might be able to rely on these and other regularities of the input to uncover crucial information about language from word boundaries over grammatical categories to sentence structure.

4. 4. Statistical Information: From Sounds to Syntax

Natural languages contain rich statistical information and children seem to have powerful learning mechanisms to access this information. The ability to track items based on their perceptual properties can allow infants to categorize their natural environment long before they have access to the meaning of words. From an evolutionary point of view it is not surprising that the infant uses this strategy in a wide array of learning tasks. Several researchers have confirmed that sequential

statistical learning occurs outside of language-learning contexts. Fiser and Aslin (2002) showed that, by mere observation of multi-element scenes, 9-month-old infants become sensitive to the underlying statistical structure of those scenes. Saffran et al. (2001) found that infants can also track the statistical features of absolute and relative pitch and intervals between these pitches in tone sequences to discover 'tone-word-boundaries'. Finally, statistical learning is not restricted to humans. Several researchers have demonstrated that cotton-top tamarins are able to track statistical cues to discover word boundaries (Ramas et al., 2000; Hauser et al., 2001), that macaque monkeys can encode and represent sequential information (Orlov et al., 2000), and that African mountain gorillas can learn sequences of complex hierarchically organized actions (Byrne & Russon, 1998). Statistical learning is a general cognitive ability that allows an organism to detect patterns in its environment across a wide range of contexts. This ability occurs throughout the animal kingdom, and it seems plausible that this domain-general ability could be recruited to support language acquisition.

The relevant question for the current inquiry is whether data-driven general-purpose learning mechanisms could account for *all* aspects of language acquisition. The empirical research conducted to date cannot answer this question conclusively. Nevertheless, there is a growing body of evidence suggesting that even complex and relatively abstract aspects of language learning that seem to require domain-specific mechanisms could be acquired using general-purpose learning mechanisms that are sensitive to statistical regularities. In the following section I provide some examples of research showing how children can exploit statistical properties of languages.

Naturally, I will begin with young infants.

4. 4.1. Sound Patterns and Semantics

One of the challenges that young learners face lies in separating individual words from the continuous stream of fluent speech. Saffran (2001) has demonstrated that young infants can use statistical cues to find word boundaries. This ability has been also confirmed for artificial nonsense languages in which additional cues (such as prosody, distinction between content and function words, etc.) have been eliminated (Saffran et al., 1996). Maye et al. (2002) have demonstrated that 6- to 8-month-old infants are sensitive to the frequency distribution of speech sounds in the input. They suggest that the ability to pay attention to the statistical distribution of speech sounds develops gradually. This ability is one crucial factor that drives the development of speech perception over the first year of life. Infants can gain access to the meaning of some words by tracking statistical regularities. They can observe that certain words reliably co-occur with certain persons or objects in their environment. The infant appears to be able to recognize co-occurrence relations long before she (explicitly) knows the meaning of words (Maye et al., 2002). The co-occurrence of names and objects seems necessary for early semantic learning. This indicates that statistical learning needs to be augmented by other cognitive processes.

Several researchers have suggested that the ability to track statistical regularities is combined with other learning mechanisms when a child acquires language (see Yang et al., 2004 for a review). Fernald (1991) has proposed that word

learning requires initially the support of prosodic, semantic, and syntactic cues. Shi and Werker (2003) have observed that the content-word preference of 6-month-old infants cannot be explained on the basis of frequency patterns (function words occur significantly more often than content words). They suggest that the infants pay close attention to the acoustic and phonological salience of content words. Similarly, Johnson and Jusczyk (2001) have shown that 8-month-old infants weigh speech cues more heavily than statistical cues. Data reported by Mattys et al. (1999) provide information about how two categories of cues (phonotactics and prosody) to English word boundaries are used and combined. Nine-month-old English-learning infants seem to have discovered that sounds that are less likely to occur together are plausible boundaries between words (known as phonotactic information), and that commonly co-occurring sounds are likely parts of the same word. These infants also show some sensitivity to the alignment of phonotactic patterns with prosodic cues to word boundaries, such as the typical placement of a stressed syllable at the beginning of a word in English. The use of these two cues in combination offers a potentially reliable segmentation tool, and it may constitute the conduit for increasingly more complex and efficient parsing strategies.

Taken together, these findings suggest that infants may integrate different cues in language learning. It might be possible that a combination of several cues reduces the complex task at hand to simpler components. Young children certainly are able to access the multiple sources of information contained in the language input (Hirsh-Pasek et al., 1996; Hirsh-Pasek & Golinkoff, 1999; Tomasello, 2003), and the combination of several general-purpose mechanisms (acoustic perception, tracking of

statistical regularities, etc.) might have the same effect as one highly complex domain-specific mechanism (such as an LAD). More research is needed to establish whether or not all the individual processes can be accounted for by general-purpose learning and whether or not the ability to combine essential information into a coherent whole requires a domain specific structure (LAD).

4. 4.2. Statistics and Syntax

The complexity of language has led several authors to conclude that it is implausible that distributional information of patterns within language could play a significant role in the acquisition of syntactic categories (Chomsky, 1975a; Crain 1991; Crain & Pietroski, 2002). However, recently, it has been shown that a considerable amount of information concerning syntactic categories can be obtained from stochastic information. In this section we introduce some findings suggesting that statistical information concerning word categories, phrase boundaries, and sentence structure can be accessed by infants and older children alike.

Generally speaking, languages that have predictive regularities support learnability because young infants are able to track these regularities (Kirby & Christiansen, 2003). Indeed, recent research has shown that combining lexical, syllabic, and acoustic cues that are contained in child-directed speech in a number of different languages (Mandarin, Turkish and English) resulted in successful classification of 80 to 90% of function words and content words (Shi et al., 1998). Redington et al. (1998) found that the distributional information contained in human

language provides a potential means of establishing initial syntactic categories (noun, verb, etc.). Finch and Chater (1991, 1994) clustered words hierarchically and obtained clusters that corresponded well with the noun, verb, and adjective categories. Similarity analyses by Monaghan et al. (2005) produced comparable results. These findings demonstrate that natural languages contain rich statistical information; this information might play an important role in the acquisition of syntactic categories.

Children learn gradually to distinguish between objects (nouns), attributes of objects (adjectives) and actions (verbs), categories that offer a way to bootstrap into grammar. In addition, it is possible that even very young children use statistical cues to draw some information from the position of words in a sentence (e.g., nouns are often followed by verbs) and from endings that indicate word-class (e.g., -ing, and -ed indicate verbs; Goodluck, 1991; Moerk, 1992; Plunkett & Marchman, 1993; Seidenberg & MacDonald, 1999; Nazzi et al., 2000b; Seidenberg et al., 2002).

Saffran et al. (1996) and Saffran (2001, 2003) have shown that domain-general learning mechanisms can exploit the transitional probabilities of co-occurring sounds to allow infants to segment individual words from the speech stream. They suggest that the same mechanisms can also account for the acquisition of abstract structure (e.g., phrase boundaries) that is not mirrored in the surface statistics of the language input in an obvious way. Their research reveals predictive dependencies in natural and artificial languages that are used by learners to locate the phrase boundaries (Saffran, 2001, 2002). According to Saffran (2003) these results indicate that learning mechanisms not specifically designed for language learning may have shaped the structure of human languages. If her claim is true, then an LAD might be

220

superfluous. Children would not need an innate language organ, but could instead rely on general-purpose learning mechanisms to track the regularities contained within language. When it can be shown that mechanisms that are clearly not innately human (such as SRNs) can extract the relevant information from natural language input, then the Poverty of the Stimulus Argument loses much of its persuasiveness. In chapter 5 I will discuss computer simulations demonstrating that it is possible to extract a wealth of information from the statistical surface patterns of the language input.

4.5. The Problem of Negative Evidence

Having discussed the positive evidence that is available to the language-learning child, I will now deal with the alleged importance of negative evidence. Broadly speaking, negative evidence provides information to learners about what is not possible in the target language. The definition of negative evidence has undergone some change during the last 50 years, and I will discuss some of the implications of this change below. Chomsky and many of his supporters have asserted over decades that negative evidence is virtually unavailable to the child. This claim in turn has been used as one more piece of supporting evidence for the necessity of an innate domain-specific LAD.

In the following section I will discuss some of the difficulties arising from incompatible attempts to define negative evidence. Next I evaluate how different definitions affect the alleged importance of negative evidence for language learning. The main focus will be on two hotly debated points. First, I will inquire whether it is

true that negative evidence is absent from the input. Second, I will inquire whether negative evidence is necessary for learning natural languages or whether it would be possible to learn a language based on positive evidence alone.

4.5.1. The Importance of Negative Evidence for Language Learning

Chomsky had stressed already in the 1950s the importance of negative evidence for learning a natural language like English. In 1967 Carl Gold published a paper that has been widely cited as 'formal proof' for the claim that a natural language cannot be learned without negative evidence (e.g., Chomsky, 1968, 1975a, 1980, 2000a, 2002, 2009a, 2010b; Baker, 1979; Marcus, 1993; Smith, 1999; McGilvray, 1999). When children begin to learn language they are faced with the task of deciding which of several mutually exclusive grammars that conform to the received input is the one of their language. Over decades it has been assumed that "[w]ithout negative evidence, children are unable to retreat from an overly general grammar to the correct limited grammar" (MacWhinney, 2004, p. 885).

If we assume that children generalize a rule of grammar (e.g., English past-tense formation) from the input they receive to all possible input, we would expect that they would form constructions like 'goed'. If we further assume that they learn from the positive evidence (input), we would expect that they would also learn 'went'. As a result, their grammar would contain both forms (goed and went). Without explicit negative evidence children could not recover from these overgeneralization errors. Because it was assumed that negative evidence is

unavailable, it was concluded that, "internal mechanisms are necessary to account for the unlearning of ungrammatical utterances" (Marcus, 1993, p. 54). However, even a cursory survey of the literature reveals that there is little agreement about what precisely constitutes negative evidence. Therefore we need to be clear about what we mean by 'negative evidence' before we can evaluate what impact it may have on language acquisition.

Initially 'negative evidence' was used to refer to two quite different phenomena. First, it was alleged that children are rarely explicitly corrected when they produce ungrammatical sentences (e.g., Chomsky, 1968, 1975a; 1986; Brown & Hanlon, 1970; Braine, 1971; Baker, 1979; Wexler & Culicover, 1980; Roeper, 1981; Berwick, 1985; Maratsos, 1986; Gordon, 1989; Pinker, 1989; Morgan & Travis, 1989; Marcus, 1993). Thus, the child will not realize that she has made a mistake but assume that her sentence was grammatical. Second, it was suggested that ungrammatical constructions, which are clearly labeled as ungrammatical, are virtually missing in the primary linguistic data. This fact would make it impossible for the child to decide whether or not a construction she has never heard is permitted by the grammar of the language she is acquiring (e.g., Chomsky, 1975a, 1980, 1986a; Baker 1979, 2003; Pinker, 1989; Marcus, 1993). In this context it was also pointed out that children are usually not presented with examples of ungrammatical constructions, which are clearly labeled as such (Marcus, 1993; Chomsky, 1988; Cowie, 1999; Baker, 2003; Lightfoot, 2005). In these cases the situation appeared straight forward: if it is impossible to learn a natural language without negative evidence and if negative evidence is not available to the language-learning child, then

natural languages cannot be learned based on the available evidence. Yet, several problems with these assumptions have been pointed out.

First, it has been suggested that the 'goal' of language learning assumed by Gold was too ambitious (Chater & Vitanyi, 2007). Gold's proof concerned ideal learners who achieve an end state of 'grammatically perfect' performance. Human learners achieve an end state that only approximates to a higher or lesser degree grammatically perfect performance. And in order to achieve this far more modest goal negative evidence is not necessary. Therefore, Gold's proof is not relevant to actual language acquisition, and under 'real world conditions' conditions, aiming at 'real world outcomes', natural languages can be learned from positive evidence alone (e.g., Feldman et al., 1969; Horning, 1969; Moerck, 1992; Cowie, 1999; Manning, 2003; Pierrehumbert, 2003; MacWhinney 2004, Chater & Vitanyi, 2007). I will return to this point in section 4.5.1.2.

Second, it has been questioned whether negative evidence is in fact inaccessible to the language learner. It seems uncontroversial that many children never encounter certain forms of negative evidence (e.g., ungrammatical constructions clearly marked as such) in their linguistic input. Children "are not given lists of ungrammatical strings...[and]... the primary linguistic data contain a proportion of ungrammatical utterances that are not flagged as such" (Cowie, 1999, p. 208). Thus children cannot expect that all ungrammatical strings they produce are clearly identified as unacceptable. However, this does not imply that children are never corrected when they make grammatical errors. MacWhinney (2004) claims that parents often provide corrective feedback that marks errors in child-speech as

ungrammatical. Sometimes caregivers provide both negative and positive feedback. Thus a mother can reject an ungrammatical utterance and/or recast the child's utterance into the corresponding correct form. While recasting is strictly speaking not negative evidence, it does provide information about parts of the child's production that are ungrammatical (for discussion see Moerk, 1992, MacWhinney, 2004).

Furthermore, other forms of negative evidence have been suggested to be potentially available (Braine, 1971; Hirsh-Pasek, Treiman, & Schneidermann, 1984; Demetras, Post & Snow, 1986; Bohannon & Stanowicz, 1988; Moerk, 1992; Farrar, 1992; Pullum, 1996; Rhode & Plaut, 1999; Sampson, 2002; Chouinard & Clark, 2003 MacWhinney, 2004). Rhode and Plaut (1999) distinguish between explicit negative evidence (feedback given to the child in response to her utterances) and implicit negative evidence (distributional properties of the input which do not depend on the language production of the child) and argue that the latter is important for language acquisition. Braine (1971) uses the term 'indirect negative evidence' suggesting that, when receiving enough input, children might be able to infer that if a certain form does not occur, then it is probably ungrammatical. This use is similar to the one suggested by MacWhinney (2004) who uses this term for differential frequency of regular and irregular verb forms in the input and shows that these differentials can have the same effect on learning as overt correction (explicit negative evidence) would have. Pullum (1996) uses the term 'implicit negative evidence' when parents/caregivers react differently to incorrect constructions (e.g., lack of comprehension, request for repetition). Similarly, several researchers have observed differential parental reaction to well- and ill- formed utterances (Hirsh-Pasek,

Treiman, & Schneidermann, 1984; Demetras et al., 1986; Nelson, 1987; Bohannon & Stanowicz, 1988; Moerk, 1992; Farrar, 1992; Sampson, 2002; Chouinard & Clark, 2003). MacWhinney (2004) shows that parents also provide implicit correction of errors by using the correct forms in response to children's errors.

Marcus (1993) calls all forms of implicit negative evidence 'noisy feedback'. He reasons it is noisy feedback because it is not given consistently from all parents and not in all instances of ungrammatical utterances of a given child (for complete list see his table 1, p.56). However, Marcus does not question the existence of negative evidence. He challenges the claim that the child receives sufficient negative evidence to learn from the input alone. Children receive neither complete negative feedback (corrective feedback given to all and only to ungrammatical utterances), nor even partial negative feedback (feedback given to some ungrammatical utterances but not to grammatical utterances). Instead, Marcus argues children receive only noisy feedback (corrective feedback given to some ungrammatical and to some grammatical utterances, p.59).

Marcus discusses several problems with noisy feedback. First, noisy feedback is so weak that a child would have to repeat an ungrammatical sentence more than 85 times to eliminate it from her grammar. Second, noisy feedback is inconsistent between parents and for specific types of errors. Third, noisy feedback is defined in such a way that it does not clearly reflect the nature of the relationship between correction of ungrammatical forms and response to grammatical forms. Thus, according to Marcus, some of the results obtained from discourse studies are methodological artifacts (p. 57). He holds that each of these three factors would

226

challenge the position that noisy feedback plays a role in language acquisition and that the combination of all three might falsify such a position. In addition some researchers have shown that even if parental feedback is available, children may not use it (e.g., McNeill, 1966; Zwickey, 1970; Braine, 1971; Pinker, 1989). Finally, noisy feedback is not available to all children.

Marcus holds that a particular type of parental feedback can only be a precondition for language learning if it (i) is available to all children, (ii) is available throughout language acquisition, (iii) is available for errors in each component of language and (iv) is available after all types or errors in each component (p. 69). Since noisy feedback does not meet these criteria, he concludes that we have "no evidence that noisy feedback is necessary for language acquisition" (p.75). Given (i) that noisy feedback is the only form of negative evidence available and (ii) that natural languages cannot be learned from positive evidence alone, we have to conclude that language learning relies on internal mechanisms that constrain the possibility space and allow recovery from errors. In the next section I will look at some replies to Marcus' arguments.

4. 5.2. Nativism and Negative Evidence

Marcus' arguments seem to lend strong support to the hypothesis that language learning relies heavily on internal mechanisms. They seem to support the

conclusion that the amount of negative evidence that is available to the child is insufficient for data-driven language acquisition. Not surprisingly, many nativists have cited Marcus' arguments in support of their view (e.g., Pinker, 1998; Smith, 1999; Lightfoot, 1999; Crain & Pietrosky, 2002). However, others have criticized these arguments (e.g., Sampson, 1997; MacWhinney, 2004). I will discuss these challenges next. First, I will introduce some points of general philosophical interest and then discuss some specific empirical criticisms of Marcus' arguments.

First, Marcus (1993) provides some compelling evidence for his claims that noisy feedback is not necessary for language acquisition. Yet, he does little to support the other crucial premise of his argument for internal language-acquisition mechanisms: that natural languages cannot be learned from positive evidence alone. In section 4.5.1.3 I will discuss challenges to this premise.

Second, Marcus stresses repeatedly that only internal mechanisms can allow the child the unlearning of grammatical errors (p. 80). From an evolutionary point of view an internal mechanism that allows the unlearning of errors seems more difficult to explain than a mechanism that avoids errors in the first place. This is because natural selection can only act on structures that are already in place. Thus, we would have to imagine that at one point humans produced grammatical errors that they could not correct. Then an internal mechanism evolved that allowed children to unlearn these ungrammatical forms. This topic is beyond the scope of this thesis, but I suggest that Marcus needs to provide some explanation for the evolutionary history of such a peculiar mechanism.

Third, Marcus' arguments presuppose that children need to learn each

individual sentence 'from scratch'. But while it is true that not all grammatical forms can be learned by analogy (see Chomsky, 1959, 1968, 1975a, 1988), it does not follow that *no* form can be learned by analogy. If at least some errors are similar to others, then the number of repetitions needed for unlearning each individual error might not be as high as Marcus suggests. For example, in overgeneralization cases the child may not need to hear a high number of sentences containing the irregular form of every new verb she learns. She could at one point discover that if she hears an irregular past-tense form, then this verb does not have a regular past tense. This eliminates the need for repetitive input.

Fourth, Marcus has nowhere shown that noisy input is not helpful at all for language acquisition. He has created a false dichotomy by presupposing that noisy input is either absolutely necessary for language acquisition or entirely useless. He has provided evidence supporting his claim that noisy input cannot be necessary for language acquisition. From that he concludes it must be useless. However, together with other forms of input even noisy input can at times provide valuable information for the language learner (for examples see MacWhinney, 2004).

Finally, like Chomsky (1986a), Marcus seems to assume the 'ideal learner' who represents all language learners (for detailed criticism of this view see Cowie, 1999; MacWhinney, 1989, 2004, Sampson, 2002, Pullum & Scholz, 2002). However, it seems far more plausible that individual children use different strategies in language acquisition. Once we remove the artificial constraint of idealized language learning and focus on actual children, we are in a better position to evaluate whether or not the available negative evidence plays a role in language acquisition. To this

task I turn in the next section.

4.5.3. Is Negative Evidence Useful for Language Acquisition?

How might negative evidence assist language acquisition? Several possibilities have been suggested. I will here only briefly touch on proposals by Cowie (1999), Sampson (2002) and MacWhinney (2004). Cowie discusses the possibility of indirect negative evidence. For example, it is possible that, when the child has predicted an ungrammatical string, she encounters a different string instead (e.g., a toddler who had expected to hear "I falled the cup off the table" hears instead "I caused the cup to fall off the table" (Cowie, 1999, p. 223)). Hearing the unexpected example can provide evidence for the child that helps her correcting her expectations. No additional form of negative evidence would be needed in this case. Cowie admits that this kind of negative evidence for a 'theory' under consideration is inconclusive; yet it is evidence that is available in principle.

Furthermore, Cowie argues that children do not learn grammar exclusively through individual sentences but by performing at least "some kind of structural analysis of incoming data" (Cowie, 1999, p. 225). For that reason every individual sentence conforming to a particular syntactic structure is confirming evidence for the hypothesis that this structure is part of the grammar being learned. On the other hand the non-occurrence of sentences that do not conform to any previously encountered syntactic structure is a type of negative evidence: it disconfirms the hypothesis that this structure is part of the grammar being learned. This type of structural analysis

provides a tool to classify some sentences that have not been encountered as negative evidence: "what makes it negative evidence is not merely its non-occurrence qua string: what makes it negative evidence is its non-occurrence qua instance of a particular syntactic structure" (Cowie, 1999, p. 225).

In addition, Cowie discusses the possibility that, while overt negative evidence (corrections of ungrammatical forms) may indeed be lacking, there are more subtle cues available that help the child to distinguish between grammatical and ungrammatical forms (Ibid., p. 229). Unlike Marcus, Cowie suggests that this type of noisy feedback can play a role in language acquisition. And she casts some doubt on the assumption that direct negative evidence is completely lacking (Ibid., p.230). In Cowie's view, children have access to several forms of negative evidence. She suggests that, while it might be correct to claim that none of these forms *alone* could be sufficient for instantaneous language-learning, a combination of several forms can assist language acquisition. She concludes, "Rather than facing a mystery of how children could learn a restrictive grammar in the total absence of negative evidence, we now confront the problem how they do in fact learn a restrictive grammar, given the kinds and amounts of negative evidence that are available to them" (Ibid., p. 234).

MacWhinney (2004) also challenges the nativist assumption that the lack of negative evidence necessitates an innate LAD. He does this by focusing on two lines of argument. First, he challenges the claim that negative evidence is not available and/or not used. Second, he argues that negative evidence is only one of many potential sources of information that can be accessed during language acquisition. Taken together, both arguments support his conclusion that the negative evidence that

231

is available to the child can supplement other sources of information and assist data-driven language acquisition.

MacWhinney challenges the proposal that parents do not correct ungrammatical utterances of their children (e.g., Pinker, 1997, Marcus, 1993). He provides several examples showing that when children produce errors "parents often provide corrective feedback that specifically marks these productions as ungrammatical" (p.887). Thus, negative evidence clearly exists. However, this feedback can have several forms (overt correction, rejection of errors, corrected recasting of utterances, implicit negative evidence), and it is not given consistently at all times or responded to by all children. This could imply that the available negative evidence is too inconsistent, too infrequent and too unreliable to overcome problems posed by too broadly construed grammars. MacWhinney is not daunted by this empirical fact but suggests that "the problem of poverty of negative evidence" (p. 895) can be overcome by a combination of different mechanisms that all *are available* to the language-learning child. Specifically, MacWhinney discusses seven 'solutions' to the problem of language acquisition (limiting the class of potential grammars, stochastic end-state criterion, conservatism, error-recovery mechanisms, competition, cue construction, monitoring, indirect negative evidence). Because space limitations do not allow discussing all 'solutions' here, I will elaborate only briefly on indirect negative evidence below. It is important, however, to keep in mind that MacWhinney proposes that none of these solutions by itself is sufficient but that "each of the solutions operates in concert with the others to support full and successful language learning" (p. 897). This is a move away from the 'one size fits

all' approach of nativists (e.g., Chomsky, 1968, 1975a, 1988; Marcus, 1993; Smith, 1999; McGilvray, 2005). Potentially children have access to all seven solutions. But for actual language acquisition the individual situation of an individual child determines which of these solutions she accesses to which degree.

MacWhinney provides one specific example to show how indirect negative evidence could assist language learning. He discusses how children could learn that 'goed' is ungrammatical without receiving overt corrections (direct negative evidence). In order to accomplish this the child needs to track the frequency of 'go' in her input (high frequency) and compare it to the frequency of 'goed' (very low frequency). By comparing these ratios to those of regular verbs ('jump/jumped', both medium frequency) the child learns that 'goed' occurs considerably less frequently than it should and is therefore ungrammatical (p. 908). In a similar manner other grammatical structures could be learned and/or errors could be 'unlearned'. Again, it important to remember that MacWhinney does not claim that this mechanism *alone* is responsible for language acquisition and/or error recovery. It is one possible mechanism available to the child, but because "it is more complicated than the basic competition mechanism and places a greater requirement on memory for tracking of non-occurrences" (Ibid., p. 909), it is likely only a 'fall back strategy' used when simple competition mechanisms have been exhausted. This is an example where negative evidence is not necessary in the strong sense requested by Marcus (1993) but where, nevertheless, negative evidence assists language acquisition.

Saxton, Houston-Price & Dawson (2005) discuss the possibility that clarification questions play the role of indirect negative evidence during language

acquisition. These authors hold that, in addition to ensuring the successful exchange of meanings between speakers, clarification questions have also grammar-correcting potential. Children at a very young age can respond appropriately to clarification questions in a range of circumstances (e.g., Gallagher, 1977; Cicognani & Zani, 1988). While many clarification questions concern semantic issues, they are also used in response to grammatical errors. Children reply usually either with repetition or revision of the utterance that prompted the question. Simple repetitions of original utterances are appropriate when the speaker assumes the listener has not heard her, but they obviously do not correct any errors contained in the utterance and are therefore not of interest here. Revisions on the other hand contain some change to the original utterance (e.g., reduction, substitution, expansion, elaboration, cuing, recasting, correction) and are of interest for "the grammar-correcting potential" hypothesized for clarifying questions (MacWhinney, 2004, p. 396). Overall specific questions elicit revisions more often than general questions (e.g., Tomasello et al., 1994) and 'looped sequences' (in which response to a clarifying question is met with another clarifying question) elicit more revisions than single clarifying questions (e.g., Brinton et al., 1986; Most, 2002). In the current context it is important that clarifying questions direct the attention of the child to the grammar of her utterance and that the revisions constitute a grammatical improvement of the original utterance. Thus from the four potential outcomes:

    (1) erroneous utterance to correct utterance

    (2) correct utterance to erroneous utterance

    (3) correct utterance to correct utterance

(4) erroneous utterance to erroneous utterance

only (1) and (3) are desirable outcomes. On the one hand we do not want the childr to respond to every clarifying question with a change to her original grammar. She should have some confidence in the grammaticality of [some of] her speech (case 3). On the other hand the grammaticality of utterances should not suffer as result of clarifying questions (case 2). This means that if error-contingent qualifying questions function as one form of negative feedback, then we should expect that they focus the child's attention on ungrammatical aspects of speech and prompt her to remember the previously learned grammatical form (c.f. prompt hypothesis, Saxton, 2000). Potentially this would make qualifying questions a valuable tool for evaluating overgeneralization errors: do children depend on innate mechanisms to recover from these kinds of errors, or do they rely on external cues such as qualifying questions. Marcus might object that not all children are exposed to the same frequency of clarifying questions (and some may not receive this type of feedback at all). The reply to this objection is that, like other previously discussed 'solutions', clarifying questions are only *one* of many possible mechanisms that can aid language learning.

To evaluate the impact of clarifying questions it is important to determine how they affect child language immediately (drawing attention to an error just made) and over time (providing additional cues for future memory retrieval). Further it is important to distinguish between questions that rely on the existence of previous grammatical knowledge and those that provide a correct alternative to the erroneous form. Saxton, Houston-Price & Dawson's results suggest that clarification requests

can aid the recall of the correct form and lead to the rejection of the child's own erroneous form. In response to error-contingent questions children replaced an erroneous form with the correct adult alternative on approximately 16–49% of occasions after the first question and even more frequently on looped sequences (MacWhinney, 2004, p. 407). However, the move from erroneous to correct forms was not the only observed result. Children also reliably resisted a change of entrenched correct forms (56-70%), and on rare occasions they replaced a correct with an erroneous form (1-2%). This shows that "clarification requests do not occasion random vacillation between grammatical and ungrammatical forms. Errors are sometimes repaired, whereas correct forms are unlikely to be abandoned" (Ibid., p. 408). The authors point out that negative feedback cannot be necessary for language acquisition, but it can support language acquisition. It highlights the possibility of a grammatical error but does not offer any information about a correct alternative. Negative evidence on the other hand contrasts child error with the correct alternative. (Ibid., p. 410). It has been shown that such contrastive modeling occurs in the input to young children and that it leads to improvements in the grammaticality of child speech (e.g., Farrar, 1992; Saxton, 1997, 2000; Saxton et al., 1998).

4.5.4. Is Negative Evidence Necessary for Language Acquisition?

In the previous section we have seen that negative evidence is available in the language input and that different forms of negative evidence can assist language acquisition. Now I will introduce some recent theoretical work showing that it is

possible to learn a natural language like English from positive evidence alone. This work challenges the assumption that natural languages are not learnable, at least in principle, without (explicit) negative evidence. For many nativists (e.g., Wexler, Culicover & Hamburger, 1975; Pinker, 1979; Hornstein & Lightfoot, 1981; Marcus, 1993; Lightfoot, 1999; Anderson & Lightfoot, 2002; Matthews, 2006), this assumption has been proven by Gold (1967). Yet, others have claimed either that Gold's proof is problematic or that it does not apply to natural language acquisition (e.g., Hornig, 1969; Feldman et al. 1969; Hirsh-Pasek & Golinkoff, 1996; Elman et al., 1996; Deacon, 1997; Cowie, 1999, Pullum & Scholz, 2002; Tomasello, 2003; Manning, 2003; MacWhinney, 2004; Scholz & Pullum, 2006; Chater & Vitanyi, 2007; Christiansen & Chater, 2008). Both sides allege that currently available empirical evidence supports their view. In my opinion this empirical evidence is inconclusive. However, recent theoretical work concerning natural language learning provides one possibility to sidestep this debate. This work cannot support the claim that children do not have language specific innate constraints but that such constraints are not *a priori* necessary for language acquisition.

Chater and Vitanyi (2007) have shown that it is in principle possible to learn a natural language like English from positive evidence alone. Their work is important because it does not rely on assumptions about the learning mechanisms available to children. Instead they introduce an 'ideal learner' who does not have access to any kind of innate language acquisition device. This means that, unlike empirical work that is always compatible with the hypothesis that children rely on an innate language acquisition device, this work can demonstrate that, in principle, it is possible to

acquire language without such constraints. On the other hand, if it is impossible for the ideal learner who has access to an unlimited amount of positive evidence to learn language, then language learning must rely on additional information, presumably in the form of innate constraints. The authors "show that there is enough information in positive linguistic data for language to be learnable, in principle, in a probabilistic sense, given sufficient linguistic data" (Chater & Vitanyi, 2007, p. 136).

Chater and Vitanyi (2007) consider a hypothetical 'ideal learner' to inquire whether or not innate constraints are logically necessary. This learner applies a Simplicity Principle, which provides the briefest representation of the available linguistic input, to the problem of language acquisition. The ideal learner is provided "with (1) the class of linguistic inputs to be learned (the linguistic 'environment'); (2) the class of possible models of the language; (3) a measure of learning performance; and (4) a learning method" (Chater & Vitanyi, 2007, p. 137). The linguistic inputs are finite strings of elements, which are generated by a Turing machine (deterministic procedure) and presented in random order. Like natural languages, this artificial language has both deterministic and random factors: "The deterministic factors determine which sentences of the language are allowed, and their probabilities—the random factors generate an actual corpus of language, by choosing a particular stream of sentences" (Ibid., p. 138). The class of possible languages is so weakly constrained that the learner is able to entertain all possible processes that might have generated the input.

Language learning is measured by how well the learner is able to predict the probability of each possible continuation of a sentence. This task is more difficult

than learning a natural language. Therefore, any learner who masters the prediction task is also able to perform the less complex task of 'learning' the language at a level comparable to that achieved by a child who learns language. In order to master the prediction task the learner searches for the "theory of the probabilistic structure of the language that provides the simplest explanation of the history of linguistic input to which the learner has been exposed" (Ibid., p. 141). Such a theory will have to capture regularities in the data to allow for accurate encoding and to distinguish between 'grammatical' and 'ungrammatical sentences'. Now the question arises whether or not under these specifications a learner who is only exposed to positive evidence can 'learn' language from the available input. Chater and Vitanyi demonstrate "that learning by simplicity can, in principle, be expected to converge to the correct conditional probabilities in predicting subsequent linguistic material" (Ibid., p. 147). These results show that if the linguistic input consists only of grammatical sentences, the learner can acquire the ability to make highly accurate grammaticality judgments and to produce grammatical language output. That is "in the asymptote, the learner can speak the language indistinguishably from the speakers in the language community in which the language was learned" (Ibid., p. 151).

Chomsky (1975a, 1985, 1986a, 2000a,b, 2002) has claimed that knowledge of certain grammatical structures could not have been acquired from positive evidence alone. To illustrate this he used sentences like (a) - (d) below. He claims that native speakers intuitively know that (a) and (b) are grammatical while (c) and (d) are not.

    (a) John is too stubborn [to talk to].
    (b) John is too stubborn [to expect [anyone to talk to]].
    (c) * John is too stubborn [to visit [anyone who talked to]].

(d) * John is too stubborn [to ask anyone [who talked to]].

However, Chater and Vitanyi suggest that the ideal learner will be able, based on positive evidence alone, to make the correct grammaticality predictions. Again, it is important to remember that the claim is not that children *are* ideal learners or even that they use strategies similar to those introduced here. Chater and Vitanyi simply demonstrate that it is not impossible, in principle, to learn how to make such intricate grammaticality distinctions based only on positive evidence. The strongly held nativist intuitions to the contrary have been wrong. These intuitions have been, in part, based on incorrect assumptions about the language-learning task. If we assume, as Gold (1967) did, that the goal of language learning is to uncover the correct theory that decides all possible cases, and is learnable from all possible texts for the language, then language cannot be learned from positive evidence alone. On the other hand if we assume that the child needs to learn from samples of the language that she actually hears to make correct grammaticality judgments and produce correct output in most cases then this goal is attainable based on positive evidence alone. Based on these findings, arguments for the necessity of negative evidence (or innate constraints) for language acquisition need to be carefully restated. Whether or not this can be done remains to be seen.

4. 6. Conclusions

In this chapter I have provided an overview of some important empirical evidence that casts doubt on the Poverty of the Stimulus Argument. It seems that the available evidence does not support many of the categorical claims put forward by proponents of the argument. Focusing on language acquisition in infancy, I explored the richness of the language input and the piecemeal fashion of early language acquisition. Research with infants has shown that many seemingly innate language-related abilities are learned over the course of several months and that this learning relies heavily on domain general abilities (such as perception). This could be an indication that the language input is proportionally more important while the requirements on innate endowment are less important than assumed by proponents of the Poverty of the stimulus argument.

Over time, children learn to integrate information from various sources into a coherent whole. Some of the strategies they use in language learning are also employed in non-language related cognitive tasks and by non-human species. In order to maintain the claim that a species-specific LAD is necessary nativists need to show that learning strategies that are shared by non-human species cannot account for all aspects of language learning.

The language input contains rich statistical information that can be used by young language learners. Infants are able to use distribution patterns of sounds to find word boundaries, and they learn to distinguish the acoustic and phonological

characteristics of content words and function words. Again, the learning mechanisms employed for these tasks seem to be general-purpose rather than domain specific.

Furthermore, I have shown that the importance of negative evidence is not as crucial as proponents of innatism suggest. On the one hand the conditions for natural language learning are different from those hypothezised for an 'ideal learner' who achieves grammatical perfect performance. The end-state of human speakers is sub-perfect, and research has shown that this state can be achieved based on positive evidence alone. On the other hand, negative evidence is not completely absent from the input. Children are exposed to explicit correction of ungrammatical forms, indirect negative evidence and noisy input (Marcus, 1993). It is correct to assert that the available evidence alone is not sufficient for language acquisition. However, together with the positive evidence it can help the child in some cases to learn important facts about language. Thus, negative evidence should not be evaluated in isolation but in the context of all available evidence.

This chapter has focused mainly on the learning processes in infants because I want to stress that learning begins very early and that children bring many acquired and not necessarily language-specific abilities to the table when they begin to tackle syntactically complex structures. A comprehensive evaluation of the LAD hypothesis needs to include all of the steps involved in language acquisition. Much research remains to be done before we know whether children who acquire language use solely the general-purpose mechanisms that are exploited in other domains of human cognition, by non-human animals or by computational models.

Chapter 5. Computational Modeling of Language Acquisition

5.1. Introduction

In this chapter I will evaluate computational models of language acquisition. The goal of this evaluation is to address the recent controversy about the usefulness of these models. On the one hand, we find the allegation that current models "make no contribution to sciences of the mind" (McGilvray, 2009, p. 24). On the other hand a growing number of researchers use computational models and claim that their findings can provide important insights regarding the mechanisms that underwrite language acquisition (e.g., Elman, 1990; 1993; Blanchard et al., 2010; Brent, 1999; Christiansen et al., 2010; Chater & Christiansen, 2008; Christiansen & Chater, 1999; Freudenthal et al., 2010; Goldwater et al., 2009; Jaroz, 2010; Monaghan & Christiansen, 2010; Rhode & Plaut, 2005; Rytting et al., 2010; Waterfall et al., 2010). I will show how models have developed over time and how newer models incorporate insights from previous models. It will become evident that several aspects of language acquisition have been successfully modeled by now. This is an important finding because the models use mechanisms that might occur in other organisms besides humans and are neither species- nor domain-specific. Further, the models suggest that language can be acquired by mechanisms that are much simpler than previously imagined. Thus, McGilvray's claim that computational work makes no contribution to the sciences of the mind should be rejected.

As discussed in chapter 4, in order to evaluate the poverty (or richness) of the

linguistic input, all sources of information contained in it need to be taken into consideration. It has become evident that statistical regularities of language provide a wealth of implicit information to the young language learner (e.g., Saffran et al., 1996; Saffran, 2002; Ellefson & Christiansen, 2000; Monaghan et al., 2005; Monaghan & Christiansen, 2008; Christiansen et al., 2010). Whether or not this information would suffice to account for all aspects of language acquisition is not known yet. One way to address this issue would be to collect comprehensive input samples from many children, analyze these samples and compare them to the output of the respective children. This is, in principle, possible, but it would be very time-consuming and expensive (Rowland et al., 2008). One solution to this difficulty is to employ computational models that simulate (at least aspects of) language acquisition (e.g., Blanchard et al., 2010; Brent, 1999; Christiansen et al., 2010; Chater & Christiansen, 2008; Christiansen & Chater, 1999; Freudenthal et al., 2010; Goldwater et al., 2009; Jaroz, 2010; Monaghan & Christiansen, 2010; Rhode & Plaut, 2005; Rytting et al., 2010; Waterfall et al., 2010). In this chapter I will give a brief overview of the history of this still-young field and discuss some recent developments. Further, I will introduce some common criticisms of computational modeling and show how researchers have responded to those criticisms.

First, it has been questioned whether or not computational models have produced any results that are relevant to language acquisition. Critics have often focused exclusively on connectionist models which are but one type of computational models: "Connectionist models of language...don't work. There are no results" (Wexler, 1991, p. 255). Over the past two decades this allegation has been repeated

virtually unchanged: "In the case of language, the evidence for connectionist models is, for the moment, about zero" (Chomsky, quoted in Smith, 1999, p. 135); "No matter how much computer power and statistics [connectionists] throw at the task [of language acquisition] it always comes out wrong" (Chomsky, 2009, p. 23) and "even the few celebrated successes [of connectionists] turn out to be failures" (McGilvray, 2009, p. 23). I will demonstrate that, while Wexler's claim might have had some justification in 1991, it is plainly false two decades later. Recent connectionist models use child-directed speech samples as input and have produced outputs that resemble closely the output of young children (e.g., Christiansen & Kirby, 2003). I will discuss some of these models in section 5.4. Furthermore, connectionist models are not the only computational models of language acquisition, and I will show that several other types of the recently employed models closely simulate the conditions under which children acquire language.

Long standing criticisms of connectionist work are (i) that connectionist models of language acquisition are not relevant to our understanding of how children acquire language and (ii) that current models omit important aspects of human language acquisition and language use. These broad criticisms deserve some consideration because to date no model exists that acquires language in exactly the same way as children do and/or acquires the full range of a human language. Given, that we currently do not know the mechanisms that allow children to acquire language, models have to focus on what is known: the input children receive and the output they produce. When a model that receives the same input as a child produces a similar output, it is possible that the mechanism used by the model could be similar to

that used by a child. However, current models are still a long distance from achieving childlike outputs. Thus, currently any inferences about language acquisition mechanisms in children are tentative.

Nevertheless, modeling has produced interesting results, and the wholesale dismissal of this work is not justified. I will show that many specific criticisms (e.g., Smith, 1999; McGilvray, 2009) are either exaggerated or based on a very superficial evaluation of computational work. I will present work that shows problems with specific claims such as that "no one finds children subjected to the training procedures for concepts or language explored by connectionists" (McGilvray, 2009, p. 23) or that connectionist models cannot possibly simulate what children do because "it takes a quarter of a million repetitions to train one to recognize *Boys like girls* as well formed" (Smith, 1999, p. 132). We have seen in chapter 4 that the input children receive before they form sentences like *Boys like girls* is much larger than nativists allow for. In section 5.4. we will see that current models do not rely on the excessive amount of input data claimed by Smith.

Researchers have become increasingly aware of the specific conditions under which children acquire language and are attempting to implement these findings in their models. I will pay special attention to models that simulate the initial steps of language acquisition (e.g. word-segmentation). These models provide important insights in the very first steps of language acquisition and suggest at least the possibility that later steps, which build on this foundation, could rely on similar mechanisms.

5.2. Connectionist Networks

Before discussing specific models I want to provide a very brief overview of computational language acquisition modeling (space limitations do not allow for detailed historical discussion. For detailed reviews see Onnis et al., 2009; Chater & Christiansen, 2008, Edelman & Waterfall, 2007; Boden, 2006). As discussed in chapter 3, Chomsky's early work on language suggested that it would be possible to provide a formally specified account of the computational mechanisms that underpin linguistic behaviour (e.g., Chomsky, 1957, 1965a). His work raised expectations that formal theories of linguistic knowledge would integrate smoothly with computational theories of language processing (e.g., parsing and production). For Chomsky knowledge of language means having (presumably implicit) access to a set of rules which in combination with computational processing operations allow the transformation from language input to language output. On this view, children acquire language by induction of a grammar that yields the set of permissible sentences of the target language. This view of natural language and Chomsky's work on formal grammars "encouraged psychologists to think about language in computational terms" (Boden, 2006, p. 298). Furthermore, based on the assumptions that cognition is modular and that language depends on domain specific mechanisms (e.g., Fodor, 1975, 1987; Fodor & Katz, 1964; Chomsky, 1965a; Schank & Abelson, 1977; Winograd, 1972; Weizenbaum, 1965; Minsky, 1968; Newell & Simon, 1976; Haugeland, 1981; Fodor & Pylyshyn, 1988), some researchers held that it would be possible to model language in isolation from other cognitive domains. This attitude

also seemed sensible given that experimental methods to investigate neuronal activity and signal transmission in the brain were fairly limited in the 1960s when the modularity approach was first developed.

Advances in neuroscience have led to a better understanding of the structure of the human brain, and connectionists have attempted to build artificial "brain-like networks" (Churchland, 1989; Elman, 1990) to simulate the processes that occur in the neurons of a human brain. In general terms, connectionist networks are modeled on brain anatomy and physiology. Connectionist networks simulate neurons, axons and dendrites and the synaptic connections between them. The networks consist of several layers of simple but highly interconnected artificial units. The sensory units at the input-layer level are directly affected by incoming environmental signals. The assembled set of simultaneous activation levels in all input units is the network's representation of the input stimulus (input vector). Input activation levels are propagated to hidden units, which sum up the received input to the hidden unit activation vector. This vector is propagated to the output layer units, which produce the output vector. Broadly speaking then the network is a device for transforming any given input-level activation vector into a uniquely corresponding output level activation vector (Churchland, 1989).

Connectionist networks are "taught" by associationist principles to recognize distinct dimensions of relevance in the presented input. The training process (back propagation of error) nudges the system towards a global error minimum. The acquired "knowledge" consists of nothing more than a carefully orchestrated set of connection weights (Churchland, 1989, p. 266). These weights are determined as a

frequency-sensitive function of the character of its training. Learning is treated as a sort of statistical modeling. Over time the network becomes sensitive to task-relevant similarities of the input patterns and can perform recognition tasks that include a large number of discriminations. Information is stored in a myriad of connection weights. The activation-vector space of a connectionist network has the capacity to embody a rich and well-structured hierarchy of categories and the network can embody representations of factors and patterns that are only partly or implicitly reflected in the input. The number of hidden units is crucial for effective generalizations. Only a system restricted by the number of hidden units is forced to find similarity relations in the presented input. We need to remember that the networks do not have any access to semantic features of the input or to rules of transformation. Instead they rely on stochastic features of the input to generate internal representations that correspond to important distinctions and structures of the input.

At first glance it may appear that McGilvray (2009) is correct to claim that connectionist networks are a modern day equivalent of Locke's *tabula rasa* and, for that reason, cannot be relevant to the language acquisition debate. In response to this criticism two lines of argument are available. First, Chomskyans hold that the available empirical evidence supports their conclusion that language acquisition depends on innate domain specific mechanisms. Thus, if it can be shown that models, which rely on internal structures very different from those of human brains, can 'learn' crucial aspects of human language, it becomes questionable that a Chomskyan LAD is necessary for language acquisition. Second, it is important to confirm (or disconfirm) *that* connectionist models are indeed the unstructured blank slate

McGilvray claims. If they are not blank slates but rely on some 'internal structure', we need to determine the origin of this structure. Specifically we need to ask if this structure was 'built in' initially or if it arose from interaction with the language input. This is an important distinction because several experimental and computational language acquisition researchers (e.g., Deacon, 1997, MacWhinney, 2004; Christiansen & Chater, 2008) have claimed that the structure contained in the input allows children to learn language. For this reason we need to pay close attention to the structure of the models that have been used by individual researchers. We will see that currently nobody works with entirely unstructured blank slates, and, when discussing individual models, I will highlight how structure is either built in or arises over time through interaction with the linguistic input. Furthermore, in addition to connectionist networks researchers have developed other types of computational models (e.g. statistical models used by Christiansen et al., 2010). Thus, even if McGilvray's criticisms would apply to connectionist networks they might not apply to all models used in language acquisition research.

5.3. Case Studies

In this section I will present several case studies that show how computational language-acquisition researchers have attempted to challenge the Chomskyan dictum that language acquisition is domain specific and depends on innate knowledge. It is important to remember that this dictum can be challenged in two different ways. First,

if it can be shown that mechanisms that clearly are not involved in human language acquisition can achieve human-like performance, then this proves that the mechanism postulated by Chomsky is not necessary for language acquisition. Second, if connectionist and/or other computational models succeed in simulating language acquisition, then this may allow inferences to the nature of the mechanisms that *are* used in human language acquisition.

It is important to remember the different implications of the two challenges. The first only refutes the strong claim that 'humans depend on an innate language specific LAD because no other language acquisition mechanism is possible'. However, even if it can be shown that another mechanism is capable of acquiring language, this has no implications for the question: 'How do humans acquire language?' Simply put, by showing that a domain independent computational device can acquire language, I have not shown that human learners do not depend on an innate domain specific LAD. The second challenge also does not eliminate this possibility. However, if it can be shown that (i) human children actually rely on the same kind of mechanisms that are used in the simulation and (ii) these mechanisms are not domain specific, then it becomes at least questionable that the mechanisms that children actually use are domain specific.

5.3.1. Elman's Early Connectionist Models

Elman (1990) has pioneered the use of connectionist models (SRNs) for language acquisition simulation. He has argued that the use of recurrent links

provides networks with a dynamic memory. "In this approach, hidden unit patterns are fed back to themselves; the internal representations which develop thus reflect task demands in the context of prior internal states" (Elman, 1990, p. 179). Elman claims that connectionist networks can 'learn' to solve relatively simple problems such as the temporal version of the exclusive 'or' (XOR) function and even 'discover' syntactic and semantic features of words. To represent temporal order in parallel-distributed language-processing models is challenging because the linguistic input is sequential. One solution to this problem is to give time a spatial representation in the model. Elman suggests that we represent time implicitly by the effect that it has on processing. By doing this we give the network a 'memory'. Usually this is accomplished in recurrent networks by adding a layer of 'hidden units' (for detailed description see Elman 1990, pp. 182-186). The hidden units allow the network to operate on its own previous output. If we add a further layer of hidden units at the input level, then we have context units, which can copy the value of the hidden unit values at an earlier time, giving the network a 'memory'. Now subsequent behaviour can be shaped by current input and previous responses. This is done "to represent time implicitly rather than explicitly [in parallel distributed processing nets]. That is, [to] represent time by the effect it has on processing and not as an additional dimension of the input" (Elman, 1990, p. 180).

To show that the networks have the potential to handle tasks requiring similar complexity as language-processing Elman tested whether they could 'learn' sequences composed of six different 6-bit binary vectors, representing speech sounds. To make the task 'language-like' these sequences were formed in two steps. First,

three consonants (b, d, g) were combined in random order to obtain a 1000-letter sequence. Then, each consonant was replaced using the rules:

b → ba

d → dii

g → guuu (Elman, 1990, p. 187)

These replacement rules produce semi-random sequences. The consonants in a series were random but the number of vowels following each consonant was predicted by the rule. After training, the networks were tested on the task of making predictions about novel input sequences, and the error rates were recorded. Unsurprisingly, it was found that the errors tend to be high when predicting consonants and low when predicting vowels. Furthermore, the network is able to make partial predictions in cases where complete predictions are impossible (e.g., based on the number of vowels it encounters, it can reliably predict that the next letter in a sequence will be a consonant, but it does not perform better than chance in predicting which consonant will occur). Finally, if there are patterns in the consonant distribution (e.g. the same sequence of consonants is repeated) performance in predicting consonants increases over time. This is because the network can make use of the 'memory' of the context units.

Using the same simple architecture, Elman showed that networks are also capable of learning more complex language-relevant structures. For example, the networks could acquire a notion functionally equivalent to 'word' as a consequence of learning the sequential structure of letter sequences that form words and sentences

but are not explicitly marked as such (Elman, 1990, pp. 191-194). And even more complex relationships are mirrored in the surface structure available to connectionist nets. The order of words in sentences reflects a number of constraints. Word order depends on several factors, such as syntactic structure, selective restrictions, subcategorization, and discourse considerations (Elman, 1990, p. 194). Using the same simple mechanisms as in previous simulations, the networks were able to 'learn' word order and simple sentence structure, based on cues available in the surface forms. This indicates that the information about alleged 'deep' structure is implicit in the surface structure of spoken language and some aspects of language can be learned based on surface structure alone (Elman, 1990, pp. 194 -203).

5.3.2. Relevance of Connectionist Simulations for the Nativist Debate

Impressive as Elman's simulations might be, we need to be clear about what they show and what they do not show. This is important because much criticism has been directed at claims that have not been made by Elman. Here I will first deal with some of the myths that are still perpetuated in the anti-connectionist literature and show that these are based on misinterpretations of the original claims and unwillingness to engage with numerous subsequent attempts to clarify Elman's position. Next, I will discuss the claims that have been put forward by connectionists and investigate whether or not they are supported by the work completed.

One of the most damaging criticisms of connectionists is that they advocate that the mind is originally a blank slate. This view has been expressed repeatedly by

Chomsky: "empiricism insists that the brain is a *tabula rasa,* empty, unstructured, uniform at least as far as cognitive structure is concerned" (Chomsky, 1977, p. 2) and there are "empiricist currents, that would have us believe that the human mind is empty, a *tabula rasa*" (Chomsky, 1980a, p. 270). That Chomsky still believes that connectionists are committed to the blank slate view is evidenced by a recent remark in the introduction to *Cartesian Linguistics* which Chomsky read and commented on "early in 2008" (McGilvray, 2009, p. 6). Apparently he did not object to this description: "[connectionists] claim that the mind is made up of 'neural nets' is innocuous; it is their claim about the initial state of the net (undifferentiated, approximating Locke's blank slate) and their view about how this net gets its 'content'...that place them firmly in the empiricist camp" (McGilvray, 2009, p. 110).

My extensive review of recent connectionist literature finds no evidence for such positions. It reveals, instead, that several researchers have explicitly or implicitly rejected completely unconstrained 'blank slate' views of language acquisition (e.g., Hare & Elman, 1995; Elman et al., 1996; Redington & Chater, 1998; MacWhinney, 2000; McDermott, 2001; Solan et al, 2005; Edelman & Waterfall, 2007; Christiansen & Chater, 2008; Chater & Christiansen, 2009). Explicit rejections of the blank slate view could indicate that early connectionist work might have had this flaw but that it has been corrected by now. Several points are important when dealing with this possibility.

First, there is an ontological point. While some connectionist researchers may initially have claimed that the networks were basically *tabulae rasae,* this was a

misunderstanding on their part. In fact, no network is ever a blank slate. There are always built-in constraints, although they may not always be recognized. These constraints can take the form of 'maximize discriminability', 'minimize processing', etc. Constraints also can be in the learning algorithm itself. Further, the structure of the network's architecture also provides a very real and powerful constraint on what can be learned (Elman, p.c.). Thus, even though connectionists and critics alike *believed* at one point that nets were *tabulae rasae*, McGilvray's claim that the initial state of connectionist nets 'approximates Locke's blank slate' never applied.

Second, in addition to the constraints that were 'built into' the connectionist networks, the models acquired additional structure through their interaction with the language input. Elman discussed this issue already in 1990. His claim was not that a completely unstructured neural net could acquire any language related structure. What he did claim was that some structure that was not initially in the network could be acquired through repeated exposure to language-like input. And, if we take Chomsky's surface/deep structure distinction seriously, accounting for the fact that the interaction with the input changes the structure of the networks is important. Yet, Elman was able to show that a similar structuring arises over time when connectionist networks are exposed to language input.

> The representations need not be 'flat,' atomistic, or unstructured. The sentence task demonstrated that sequential inputs may give rise to internal representations which are hierarchical in nature. The *hierarchy is implicit in the similarity structure of the hidden unit activations* and does not require an a priori architectural commitment to the depth or form of the hierarchy. Importantly, distributed representations make available a space which can be richly structured (Elman, 1990, p.208, emphasis added).

What Elman says here is not that nets are initially completely unstructured blank slates but that they do not need to have one specific fixed (innate) structure in order to solve the sentence task. And this, he suggests, is the case because in the case of language much of the structure is contained in the input: "What is exciting about the present results is that they suggest that the inductive power of the PDP approach can be used to *discover structure* and representations in tasks which unfold over time" (Elman, 1990, p. 209, emphasis added). The language acquisition task undoubtedly is one that unfolds over time. For Elman the connectionist models can help uncovering how much of the structure contained in language output can be acquired from the structure of the input. But this does not entail that just *any* network could succeed in this task. Other researchers have stressed the same point:

> [models] provide insight into which aspects of network performance are due to architectural biases and which arise due to learning. A network always has some bias with respect to a particular task, and this bias is dependent on a number of factors, such as overall network configuration, the nature of the activation function(s), the properties of the input/output representations, the initial weight setting, etc. (Christiansen & Chater, 1999, p. 195)

Thus, the simulations will also help us to discover how much structure needs to be 'built in' to the networks and whether or not this structure needs to be task specific. This leads to a second common criticism of connectionist models. It is often alleged that general-purpose learning mechanisms are, in principle, not able to solve the language- learning task (e.g., Chomsky, 1959, 1966, 1975a, 1986a, 2005, 2007, 2009a; Marcus, 1993; Smith, 1999; Crain & Pietroski, 2002; McGilvray, 2005, 2009).

This criticism was initially leveled against behaviourism. But over the years it has become a criticism of any language-acquisition account that does not posit innate domain-specific knowledge and/or mechanisms. I will discuss here briefly the original criticism of behaviourism and then show that, even though 'empiricist research strategies' have changed considerably since the 1950/60s, the core elements of current Chomskyan criticisms remain the same.

Chomsky's (1959) review of Skinner's *Verbal Behaviour* is often considered as the 'death blow' to behaviourism. A detailed discussion of this review and its implications would lead us too far afield (for overviews see Harris, 1993; Boden, 2006), but some of the fundamental assumptions that motivated Chomsky's criticism are relevant here. Essentially Chomsky alleges that behaviourists hold that *all* knowledge is result of stimulus-response conditioning and deny that innate knowledge and/or 'internal mechanisms' play a role in language acquisition. Chomsky claims, "For Skinner, what we call 'knowledge of French' is a 'repertoire acquired as a person learns to speak French'" (Chomsky, 1971, p.7). Obviously, these repertoires are always finite and do not contain countless legitimate sentences of the language that is learned. Chomsky asks rhetorically, "But what does it mean to say that some sentence of English that I have never heard or produced belongs to my 'repertoire,' but not any sentence of Chinese?" (Ibid.) To stress how severe the problem is Chomsky continues, "Skinnerians, at this point in the discussion, appeal to 'similarity' or 'generalization,' but always without characterizing precisely the ways in which a new sentence is 'similar' to familiar examples or 'generalized' from them" (Ibid.). Thus, the predictable answer to the rhetorical question is that it is nonsensical

to expect that *any* sentence that has never been heard before could be 'recognized' based on similarity or generalization alone. But is that really the case?

In order to evaluate Chomsky's claim we would need to know how 'different' from previously encountered sentences a never-heard-before sentence would need to be. Can a child who has heard (1) "This is a blue ball" and (2) "This is a yellow cup" know that the 'never-heard-before' sentence (3) "This is a blue cup" is an English and not a Chinese sentence? There seems to be no *a-priori* reason to assume she could not. But possibly (3) would not qualify as a 'never-heard-before sentence' because all words in (3) have been heard before.

Here we encounter again a by now familiar problem with Chomsky's science. While Chomsky criticizes 'Skinnerians' for their vague terminology, he does not give the necessary and sufficient criteria that a sentence must fulfill to qualify as 'never-heard-before' and how we would verify that a given sentence has in fact never been heard before. Without ever citing any evidence for this claim, Chomsky expresses his conviction that the majority of the sentences we understand and produce "bear *no physical resemblance* to sentences which are 'familiar'" (Chomsky, 1966, p. 11, emphasis added). The term 'physical resemblance' is never explained, leaving it for the reader to decide whether or not two particular sentences bear *any* physical resemblance. Most people probably will agree that (3) bears *some* resemblance to either (1) or (2). But consider the following examples:

(4) The moon looks like a ball.

(5) The sun is hot.

(6) My birthday was last week.

Arguably (4) bears *some* resemblance to (1) because both sentences contain the word 'ball'. (1), (2), (3) and (5) all contain the word 'is'. On the other hand (6) does not contain any of the words contained in the other five sentences. But does this imply that this sentence bears *no physical resemblance* to any of the other sentences? As I have shown in section 4.3.1.2, a neonate is already able to discriminate sound structure differences of languages that belong to different rhythmic classes (Nazzi, Bertoncini and Mehler, 1998). We also know that it takes the infant several months to learn also to distinguish her own language from other languages of the same rhythmic class (Nazzi, Jusczyk and Johnson, 2000). It seems reasonable to assume that this learning process is based on *some* physical resemblance that the sentences of one language bear to each other but not to sentences of a different language. Thus, the question of how I can know that a sentence I have never heard before is English and not Chinese may indeed have a surprisingly Skinnerian answer: Because some of the physical properties (sound structure etc.) of an English sentence resemble those of other English sentences while the physical properties of a Chinese sentence do not.

At this point it might be claimed that taking Chomsky's rhetorical question literally is not only uncharitable but also misses the more subtle point he is making. To this I respond in two ways. First, the onus is on Chomsky to express his position in an unambiguous way. Rhetoric has always played an important role in his 'science', be it to attack opponents or to bolster his own position. On the one hand, he accuses opponents of vagueness and imprecision: "the Skinnerian must lapse into

260

mysticism (unexplained 'similarities' and 'generalization' of a sort that cannot be specified) as soon as the discussion touches the world of fact" (Chomsky, 1971, p.16). On the other hand, his own arguments are frequently vague and imprecise (for detailed discussion see Botha, 1986; Sampson, 2002). Thus, exposing the limitations of his rhetorical arguments is not just an idle waste of ink.

Second, even on a more charitable reading Chomsky's argument that 'empiricists' have never moved past behaviourism is no longer supported by evidence. Below is a recent reformulation of his criticism:

> ...people argue that environmental factors are critical but without offering any account of the facts in question in terms of such alleged factors. And as long as they don't produce any moderately plausible account in terms of presumed environmental factors, all I can say is that they're not holding my attention. It is not very interesting if somebody claims that something is the result of the environment or an act of God or electrical storms in the vicinity, or whatever, if they don't provide some explanatory scheme that can at least be investigated. (Chomsky, 1993, p. 4)

As we can see, the criticism has remained virtually unchanged, as has the rhetoric in which it is couched. Similar accusations regarding lack of clearly specified learning procedures and thinly veiled contempt for the work of those who defend different views are expressed in the following passage by McGilvray:

> ... empiricists seem to have added little to Locke. And like Locke's efforts theirs generally fail to meet the conditions on adequacy of a *naturalistic* theory...they stipulate that what is in the head must be some kind of neural net that has the property that it can be modified by 'experience' and correction of its outputs... To show the merits of their 'hypothesis' (which is rarely if ever explicitly stated), they might introduce computer models in the form of 'neural nets'... They consider their efforts successful if the computer model 'learns' to perform the 'task' set it to the satisfaction of the 'experimenter'. This story makes language acquisition

a miracle...(McGlivray, 2009, pp. 20 - 21, original emphasis)

Now returning to Elman's work, we have to ask whether he commits the sins Chomsky (1993) and McGilvray (2009) allege. Undoubtedly Elman (1990, 1993, 1999) has attempted to show that the language input (=environment) contains information that is relevant and important for language acquisition. The input does to some degree determine the output. However, Elman also has shown that the relationship between input and output is not merely one of stimulus-response. If this were the case, the connectionist nets would fail in any tasks that require dealing with previously unencountered examples. Elman claims that some of the information that allows the nets to deal with new examples is contained in the input. However, nowhere does he claim that *all* relevant information is contained in the 'environment' and he acknowledges that the information contained in the input alone is insufficient. Specifically, he says:

> While it is undoubtedly true that the surface order of words does not provide the most insightful basis for generalizations about word order, it is also true that from the point of view of the listener, the surface order is the only visible (or audible) part. Whatever the abstract underlying structure be, it is cued by the surface forms, and therefore, that structure is implicit in them. (Elman, 1990, p. 195)

Here Elman suggests that it might be possible to use regularities "on the surface" (that part of language that is available as input) to uncover regularities in the 'abstract underlying structure'. It is of course possible that this approach will turn out to be wrong. However, it is not the case that Elman does not "produce any moderately

plausible account in terms of presumed environmental factors" (Chomsky, 1993, p. 14). Quite to the contrary, he offers a detailed account of the input (= presumed environmental factor) he uses for his simulations (Elman, 1990, pp. 187-188, 193, 195-196, 200). Furthermore, it is not true that he appeals vaguely "to 'similarity' or 'generalization,' but always without characterizing precisely the ways in which a new sentence is 'similar' to familiar examples or 'generalized' from them" (Chomsky, 1971, p.48). Quite to the contrary, Elman carefully specifies in which ways the novel sentences are similar to those the nets encountered during 'training' (Elman, 1990, pp. 195 - 197). He outlines details regarding the experimental procedure, the input, the expected output and the actual performance of the nets.

Elman continued his work on SRNs and explored, for example, the learning of sentences with embedded clauses (Elman, 1991), and complex embedded structures (Elman, 1993). This work showed that tasks, traditionally thought to require an explicitly recursive computational structure, could be solved by the simple network architecture of an SRN. Here Elman also attempted to implement the insight that in humans "learning and development interact in an important and non-obvious way. Maturational changes may provide the enabling conditions which allow learning to be most effective" (Elman, 1993, p. 72). He demonstrated that in some circumstances, SRNs that are trained to represent part/whole relationships and embedded clauses of complex sentences "work best when they are forced to 'start small' and to undergo a developmental change which resembles the increase in working memory which also occurs over time in children" (Ibid.). In other cases, 'learning' can only occur when the entire data set is available to a network (e.g., Harris, 1991). Elman claims that

"the deeper principles which underlie learning in the general class of connectionist systems which rely on error-driven gradient descent techniques... interact with characteristics of human development in a beneficial manner" (Elman, 1993, p. 72). Given that infants also start with limited memory capacity and only pay attention to a small segment to the linguistic input they receive, connectionist networks simulate one important aspect of human learning in general and language acquisition in particular.

Elman's approach has been critiqued by Rohde & Plaut (2003), who suggest that artificial languages that only contain the relevant syntactic information are not a good representation of human languages. Human subjects rely to a considerable degree on semantic information when processing sentences that contain long-distance dependencies and it is questionable that SRNs that are deprived of access to semantic information perform a task that is sufficiently similar to the task faced by human children. For this reason Rohde and Plaut performed experiments using an artificial language that provided syntactic as well as semantic information (for details see Rohde & Plaut, 2003, pp. 2-4). Unlike Elman they found that networks that were exposed to complex constructions throughout training outperformed those that 'started small'. "Under no condition did the simple training regimen outperform the complex training" (Ibid., p. 5). These authors believe "that recurrent connectionist networks already have an inherent tendency to extract simple regularities first" (Ibid., p. 18). The network first learns short-range dependencies and considers long-range constraints as noise. Once the short-range dependencies are learned, the network can use the available information to learn long-distance dependencies. This is an 'innate'

constraint, but it is not domain specific. Similarly, in children the specific stages of language acquisition could be caused by a cognitive "system that is unorganized and inexperienced but possesses great flexibility and potential for future adaptation, growth and specialization" (Ibid., p. 21).

The SRN simulations discussed so far focus on the acquisition of syntactic structure, which is just a small part of the overall language-learning process. The promising results do not imply that all aspects of language acquisition can be modeled by connectionist nets. But they challenge the claim that "there is no reason to take empiricist speculations at all seriously ... [because] the apparent aim is not to explain facts of human language and concepts and their growth" (McGilvray, 2009, p. 23). If it can be shown that the 'learning' in connectionist nets resembles that of human children in important aspects, then this work should be taken seriously. And as long as we do not know the mechanisms that underlie language acquisition in human brains, we cannot rule out that they are similar to those used in connectionist models.

Obviously, the method and results of connectionist researchers can be critiqued. Yet, Chomskyans rarely offer targeted criticism of specific work. Instead they offer blanket condemnation: "... there is no reason to take empiricist speculations seriously at all... dogma not reason drives the empiricist research strategy" (McGilvray, 2009, p. 23) and conclude that "empiricist efforts like these make no contribution to sciences of the mind" (McGilvray, 2009, p.24). As the discussion of some of the work of Elman and other computational language researchers has shown, it appears at least somewhat implausible that dogma alone drives their research strategy. Presumably they would not perform time-consuming experiments if they

were convinced that their view cannot be mistaken. And, regardless of whether or not the simulations bear any resemblance to processes that occur in human language acquisition, the findings of connectionists can help to determine the path of further research. For example, this research can help to determine whether or not we need to postulate complex internal representations in order to account for language acquisition and processing. Using a simulation that did not contain such representations was an exploratory step with not necessarily expected results: "The approach described here employs a simple architecture, but is surprisingly powerful" (Elman, 1990, p. 207).

The final criticism I want to introduce here suggests that connectionist work is not relevant to language acquisition because the models rely on mechanisms that are very different from those employed by children. Again, this criticism needs to be unpacked. First, there is the claim that "[n]o one finds children subjected to the training procedures for concepts or language explored by connectionists" (McGilvray, 2009, p. 24). This would imply that it might be possible that connectionist nets can be 'taught' to perform in a language-like manner. However, because their performance is the result of learning strategies that are different from the strategies used by children, their 'accomplishment' is not relevant to language acquisition in children.

Three points are of importance here. First, the criticism was made in 2009 and aimed at 'connectionists'. This means that McGilvray must hold that all connectionists currently use 'training procedures' that are not relevant to human language acquisition. In section 5.3.3. I will discuss how connectionists models have changed over the last 20 years and evaluate whether McGilvray's criticism still

applies to all models.

Second, McGilvray seems to imply that connectionist nets actually do achieve performance that resembles language acquisition. If so and if he is also correct in holding that the mechanisms connectionists explore are different from those used by children, then some mechanisms not used by children can result in language acquisition. In other words, it is not necessary that language acquisition relies on a species-specific innate mechanism. This would refute one of Chomsky's dicta that only mechanisms that are specifically dedicated to language can deal with the task of language acquisition (e.g., Chomsky, 1966; 1975a, 1985, 2005, 2007). Presumably this threat to the Chomskyan program justifies not only the content but also the hostile tone of McGilvray's critique of the 'connectionist work':

> [empiricists hold that] there must be a lot of the mind/brain that is plastic and can be modified by experience and training procedures, for that is where language must go. To show the merits of their 'hypothesis' (which is rarely, if ever, explicitly stated), they might introduce computer models in the form of 'neural nets' of what they believe are the plastic regions of the mind/brain and subject these computer models to input that is supposed to simulate their views of human linguistic experience and the data of language learning. They consider their models successful if the computer model 'learns' to perform the 'task' set it to the satisfaction of the 'experimenter'. (McGilvray, 2009, p. 21)

McGilvray is convinced that this approach must fail. Yet, his indiscriminate lumping of opposing views under the headers 'empiricist' or 'connectionist' makes his critique vulnerable to counter-examples. As we will see in section 5.3.3, the input used in current models of language acquisition is based on natural language samples recorded from actual input that real children receive. Whether or not the connectionist models

resemble how 'plastic regions' of the brain are affected by language input is certainly an open question. But, again, an important part of Chomsky's argument is the claim that language learning can only occur because children have an innate domain-specific language acquisition device. Thus, if a mechanism that does *not* resemble human brain regions can acquire language, the necessity of an innate domain-specific mechanism can no longer be assumed.

Third, we need to ask if McGilvray is correct to hold that "massive training procedures" (p. 23) are irrelevant to language acquisition. They certainly would be if children acquire language, as claimed by Chomsky, (nearly) instantaneously (Chomsky, 1975a, 1980a) "without effort" (Chomsky, 1985) in a way that is "virtually uniform" (Chomsky, 1986a). According to Chomskyans, all normal children display an "effortless creative use of language" McGilvray, 2009, p. 17) at an early age. But in chapter 4 we have seen that language learning is not as effortless as claimed by Chomskyans.

We encounter again one of the false dichotomies: either children acquire language by a cartoon-like 'empiricist' learning strategy, or they must rely on a domain-specific innate language-acquisition device. No middle-ground is explored. We have the implausible scenario that "much of the child's early life and use of language is devoted to focused data-gathering and training sessions that consist in getting the child to conform to the 'speech habits...or uses of concepts the child's trainers want it to exhibit" (McGilvray, 2009, p. 16). Furthermore, the cartoon-empiricist is accused of misrepresenting the social practices: "Where are all the trainers the learning procedure requires, and even assuming that there are some, where do they find the time?" (McGilvray, 2009, p. 17). If we accept that children do not spend most of their day gathering data and parents do not spend most

of their time 'training' their kids, we are forced to admit that language learning cannot depend on mechanisms suggested by connectionists. At least that seems to be what McGilvray wishes to imply. But is the amount of input data required by connectionist models not available to children? As discussed in chapter 4, the average toddler has heard between 7 and 10 million utterances. And while much of the input may not have the structure of explicit teaching or feedback and error correction, it is not clear that this structure would be needed for language acquisition. Connectionist models hope to offer a snapshot of some aspects of early language acquisition. Once a child has learned a certain number of concepts, it might be possible that the amount of input she requires to learn new concepts is substantially reduced.

So far I have discussed models that do not simulate what children are doing (e.g., Elman, 1990, 1993). This work was also not based on input actually received by children. Thus, these models, while allowing us to draw some conclusions regarding the ability of simple mechanisms to acquire language relevant 'knowledge', were not 'replicating' what is going on in human language acquisition. In the next section I will discuss models that use child-directed language as input and better address the challenge that modeling has nothing to do with the conditions under which children acquire language. McGilvray (2009) does not discuss any specific connectionist models. This makes it difficult to judge whether or not his criticism is justified. Assuming that we can pick any connectionist model and find it problematic in the ways discussed by McGilvray, I will look in the next section at some specific models. Here I shall keep in mind that we need precise models "that can be tested, refined or rejected on the basis of publicly available data and/or replicable experiments" (MacWhinney, 2010, p. 477).

5.3.3. Other computational models of language-acquisition

Over the last 20 years researchers have refined computational models and many of these models rely on "corpora of spontaneous adult–child interactions made available through the Child Language Data Exchange System (CHILDES)" (MacWhinney, 2010, p. 477). These CHILDES corpora consist of recorded samplings of adult speech that serves as input to language-learning children. This practice refutes McGilvray's claim that modeling is based on arbitrary pairings of "input to outputs that suit the experimenters' criteria for correct behaviour" (McGilvray, 2009, p. 22-23) and has nothing to do with the conditions to which real children are exposed. MacWhinney suggests that "the job of the computational modeler is to determine a set of algorithms that can take the child-directed speech (CDS) as input and produce the learner's output (LO) at successive developmental levels" (MacWhinney, 2010, p. 477). This means the input for the models resembles closely the input children receive, and the models are informed by the developmental stages that are typical for children.

I will now take a closer look at some of the recent 'empiricist models' and assess how closely they simulate human language acquisition. As we have seen in chapter 4, children have to acquire a complex set of skills long before they are able to comprehend and produce complex grammatical structures. Recent computational models attempt to simulate several of the important steps that children take on the road to language. It would lead too far afield to discuss all relevant models here, so I will focus on some of the important milestones.

5.3.3.1. Models of Speech Segmentation

Recently researchers have attempted to use computational models to simulate speech segmentation acquisition. As discussed in section 4.3.1.2, infants need to learn to segment the continuous stream of language input into individual words. They master this skill in the course of several months. Potentially there are many ways to achieve speech segmentation. But only modeling some of them will give us a better understanding of how children might accomplish this task. Researchers are aware of this and have tested a variety of them by now: "Previous developmental models of speech segmentation differ substantially across a number of parameters, including whether the model builds a lexicon, segments words by clustering smaller units or breaking down larger units, or incorporates external constraints on performance" (Monaghan & Christiansen, 2010, p. 546). The performance of these models on essential criteria differs. For example, some models achieve a high degree of precision (correct identification of words from the input) but rely on mechanisms that are not psycholinguistically plausible (e.g., access to the complete input, no memory limitations, optimal learning). However, even psycholinguistically implausible models can provide us with a better understanding about the information content of the input. If it turns out that the information present in the input alone is sufficient for speech segmentation, then this task does not necessarily depend on innate domain specific 'knowledge'.

Like many researchers, Monaghan and Christiansen assume that the information contained in the input is helpful in the acquisition of speech segmentation. However, they

are not only looking for a mechanism that can extract the relevant information from the input but are also attempting to provide a model (PUDDLE) that closely resembles how children accomplish this task. As input for this model they use speech spoken by adults in the presence of children aged 2 years, 6 month or younger (six English CDS corpora from the CHILDES database (MacWhinney, 2000); for complete details see Monaghan & Christiansen, 2010, pp. 552-554). This input for the model is similar to input received by children who learn language.

PUDDLE is similar to young children is several aspects. Like children, the model builds its lexicon incrementally from the input. This 'strategy' does not require that the model makes multiple, simultaneous decisions about the match between a given utterance and the acquired lexicon. Just like young children, the model is initially unable to perform complex cognitive tasks simultaneously. PUDDLE simulates how children can take advantage of features that are readily accessible in CDS and can accommodate learning. The model performs like a child because "the memory resources and computational requirements are minimal" (Monaghan & Christiansen, 2010, p. 248)

Monaghan and Christiansen were especially interested in two of the readily accessible features of the input: (i) utterance boundaries and (ii) the interspersal of high frequency words in speech (Ibid.). As the results show, these two cues in combination can go a long way towards correct speech segmentation. This is because CDS contains a relatively high percentage of single word utterances (26% in the CHILDES corpus), and some of these words occur frequently (e.g., the child's name). Treating utterances as words and recording the frequency of words allows PUDDLE to build up a lexicon incrementally, and items from the lexicon are in turn used to determine 'new' words (e.g.,

parts of utterances that precede or follow an item already in the lexicon). It was possible to show that a small set of frequently occurring words can help in "carving up the rest of the speech stream into its constituent words" (see Monaghan & Christiansen, 2010, p. 250, for full details). The performance of this model, which relies on a very simple algorithm for discovering words, was impressive. Depending on the corpus used, the recall was 70-79%, meaning that PUDDLE identified between 70-79% of the words that were contained in the input. Precision was 70-76%, meaning that 70-76% of the words identified by PUDDLE were indeed words from the input. This may sound like substantially less recall and precision than children might achieve. However, we need to keep in mind that the model was only exposed to 10.000 utterances in total and that it did not have access to numerous other cues that are used by children (e.g., acoustic, phonological and prosodic cues; for overview see Monagahan & Christiansen, 2008). These results suggest that all the information contained in the input might be sufficiently rich for speech segmentation and that this information might be extracted with relatively simple mechanisms.

Before discussing models that simulate other aspects of language acquisition, I want to evaluate whether PUDDLE is superior to Elman's connectionist networks in addressing the main criticisms of computational modeling. To reiterate, critics of this work suggest that two aspects need to be considered to assess whether or not a model is relevant to language acquisition in children. First, the input used for modeling must be comparable to the input children typically receive. Second, the mechanisms used by the model must be similar to those used by children. Intuitively it seems that using corpora from the CHILDES database refutes the claim that empiricist modeling has "nothing to do with the conditions under which children acquire ... language" (McGilvray, 2009, p. 23).

If the input used for the model is the same as input received by actual children, modeling has at least something to do with the conditions under which children acquire language. Thus, PUDDLE fares better than Elman nets, addressing the first point of criticism.

Still, a further criticism suggests that the fact that the conditions under which the model 'learns' speech segmentation have something to do with those under which children learn does not guarantee that the learning conditions are sufficiently similar. In particular, the practice of using quantitatively limited samples of CDS has been critiqued:

> The samples are limited. What is recorded varies from a few hours in total to an hour a week for several years. Only some of it is natural conversation, and only some is an adult speaking. What adults (or the television) said when the tape recorder wasn't running will never be known, so this is not a good method for tracking lacunae. (Fodor & Crowther, 2002, p. 108)

The overall point of the critique seems to be that the CHILDES samples are only a small segment of the input that children actually receive. This is a fair criticism if these samples are used to draw conclusions about the complete input that children receive. However, in the context of computational modeling the limited CHILDES samples can still be useful for several reasons. First, one claim of the Poverty of the Stimulus Argument is that a child could not have learned a rarely occurring syntactic rule from the complete input she received. When it can be shown that a model can 'learn' this syntactic rule from input that is only a small subset of the actual input, then it should be possible that this syntactic rule also can be learned from the complete input. Second, another common criticism of empiricist models is that they are based on input that was available to privileged children. Several of the initial

CHILDES transcripts were obtained from the input children of language acquisition researchers received (e.g., Brown, 1973). However, all children acquire language even those who receive considerably less input than 'the average' learner (Chomsky, 1985, 2007). So again, it would seem to be an advantage that even a model that relies on quantitatively limited input can 'learn' the syntactic rule from this input. In these two cases, then, it seems to be an advantage to use samples that are only small subsets of the actual input. Models that are able to learn features that have been claimed 'unlearnable' from these limited samples should also be able to learn these features from the complete input.

Nevertheless, the criticism that some computational models rely on a limited sample size is not entirely unjustified. It might be possible that the input used in modeling contains fewer incorrect or incomplete utterances ('noise') than the input children typically receive. If this were the case, then the language learning task for the models might be considerably easier than that faced by children. To overcome this potential problem it needs to be determined if the corpora available through CHILDES are representative samples of input. Should this not be the case, several remedies are available. First, the relationship of grammatical utterances and 'noise' in the input needs to be determined. If a high percentage of the input consists of grammatical utterances, it may be acceptable to test only small samples. If the samples contain a high percentage of noise, the sample size needs to be increased. Second, it would be possible to insert some noise into existing corpora and test if the performance of models deteriorates. If this is the case, the models need to be adjusted to deal with more realistic input. Third, more complete input samples could be

collected and used. Given the problem of time and cost (e.g., Behrens, 2009; Rowland et al., 2009) this may not be a feasible solution. However, it is in principle possible. Thus, there is no *a priori* reason to suspect that the potential problems of using only partial input for models like PUDDLE are insurmountable.

Another common criticism of modeling is that there is not enough acoustic variability in the input used for modeling. Children have to succeed with the word-segmentation task after exposure to input that varies between speakers and even between different utterances of the same speaker (Newman et al., 2006; Newman, 2005; Brent & Siskind, 2001; for proposed solution see Christiansen & Allen, 1997). While competent speakers can rely on context and other cues to disambiguate and/or recognize unclearly or mispronounced words, infants lack this knowledge. For modeling, the CHILDES data are transcribed, and in the process incorrect utterances, mispronounced words, etc. are eliminated. It has been argued that models relying on this 'sanitized' input do not provide a realistic simulation of language learning (McGilvray, 2009; Chomsky, 2007; Smith 1999). To address this challenge Rytting et al. (2010) developed a method to test whether models can deal with probabilistic, speech-derived input which "is more realistic than the types of transcripts usually used to train and test computational models of word segmentation" (Ibid., p. 514). They used as input "recordings of four mothers... directed at infants age 0;9 to 0;10.26" (Ibid., p. 525). From this input they removed 'problematic utterances' such as "whispered or sung speech; unintelligible, untranscribed or partial words; word play or pet names... [leaving] 13,443 utterances for the four mothers" (Ibid.) but left utterances that were clearly audible yet grammatically incorrect or incomplete. They

found that the performance of SRNs and models used by other researchers (e.g., Christiansen et al. 1998) "is robust for data with subsegmental variation when this variation is carefully controlled" (Ibid., p. 530) but that an increase in variation leads to a significant degradation of the performance.

Rytting et al. hypothesized that increased variability in the input "compromises the reliability of the segmental cues, such that it is no longer possible to find word boundaries using these cues alone" (Ibid., p. 531). The authors then tested the impact of additional cues (e.g., dictionary-derived stress cues and hyperarticulation of word-initial syllable cues) and found that when the models are "faced with highly variable, potentially ambiguous input, multiple probabilistic cues still outperform each cue separately" (Ibid., p. 536). These findings indicate that natural language contains an abundance of cues for word-segmentation and that the combination of several of these cues makes the segmentation task easier even under conditions that are less than ideal. Furthermore, infants also might be "able to detect regions of clear speech, and treat the beginnings of such regions as likely word boundaries" (Ibid., p. 540).

These findings indicate that noise in the input can interfere with the performance of simple models that use a small training corpus and rely only on a limited number of cues. This underscores the need for models that have access to a variety of cues (e.g., Monaghan & Christiansen, 2009). However, simpler models can be helpful in establishing which cues are exploited at which stage of language acquisition. And, in ruling out possibilities that were not *a priori* recognizable as 'dead ends', even unsuccessful models contribute to our better understanding of the

mechanisms that allow for language acquisition.

5. 3.3.2. Multiple Cue Integration Models

Research with children has shown that language contains multiple statistical cues and that children are able to access this information (e.g., Monaghan & Christiansen, 2009). Having access to more than one source of information can simplify the language-learning task because children learn over time to integrate the information from several cues into a coherent whole. Given that multiple cue integration assists children in language learning, it is desirable to simulate this process in computational modeling. It would seem that models that can access the multiple sources of statistical information that are contained in the language input will out-compete models that rely only on one source of information. In addition, because those models simulate more closely how children acquire language, they will also address the Chomskyan challenge that models "have nothing to do with the conditions under which children acquire language" (McGilvray, 2009, p. 23).

Recently many researchers have begun to test whether it is possible to simulate the effects of multiple cue integration with computational models (for overview see Monaghan & Christiansen, 2009). The challenge is to design models that can access simultaneously several cues and combine the information to assist word-segmentation. I will discuss here one of these models. Blanchard et al. (2010) propose that infants can learn individual words based on frequent occurrence (e.g., their name, 'mom', frequent function words) and/or language specific phonotactic

constraints (stress patterns, allophonetic variation, etc. as discussed in chapter 4). According to Blanchard et al., frequently occurring words form the first tiny lexicon, which allows the learner to infer some phonotactic constraints. This information in turn can help to recognize additional words. This combination of these two cues solves one important problem that beginning language learners face: how can they know which phonotactic constraints apply before they know words and vise versa. Thus, "knowledge of familiar words, combined with increasingly refined phonotactic constraints, support and reinforce each other in speech segmentation" (Blanchard et al., 2010, p. 491).

Blanchard et al.'s model PHOCUS relies on very basic assumptions about language learning. Beginning with an empty lexicon, it incrementally adds items to the lexicon, based on phonemes that occur together (probabilistic and phonotactic cues). As discussed in chapter 4, phonemes that occur within frequent words have high transitional probabilities while phonemes that cross word boundaries have low transitional probabilities (e.g. Saffran et al., 1996). In addition to these transitional probabilities, which are helpful to detect word boundaries, Blanchard et al.'s model could exploit phonotactic cues. Specifically, when the model encountered an unfamiliar word, it could rely on two kinds of phonotactic cues (phoneme combinations and occurrence of at least one syllabic sound per word; for details see Ibid., pp. 496 - 501).

Blanchard et al. could show that the combination of these two simple cues allowed a performance of 76-81% precision/recall scores for an English test corpus. Unexpectedly, the same model performed substantially worse (19 - 47%

precision/recall scores) on a Sesotho corpus (Ibid,. p. 503). This result "highlights the importance of testing acquisition models on data from a variety of languages because the results can be so different from what is obtained with English corpora" (Ibid., p.505). The authors explain this difference in performance with the fact that the most frequent word in the Sesotho sample is monosyllabic. This results in a very high percentage of over-segmentation errors from which the model cannot recover. Obviously, children learning Sesotho are able to master the word segmentation task. This indicates that they cannot rely on the same cues as the model used by Blanchard et al.. A model that incorporates more cues than the Blanchard et al. model simultaneously might succeed in the Sesotho word segmentation task. What these cues are and how they interact with one another is a matter of ongoing research.

Taken together, the results from modeling word-segmentation support the following points relevant to my inquiry. First, the models I discussed did not confirm the challenge that computational modeling has "nothing to do with the conditions under which children acquire ... language" (McGilvray, 2009, p. 23). It remains possible that such models exist, but the onus is on McGilvray to identify them and demonstrate that they commit the sins he alleges.

Second, computational modeling in general and connectionist models in particular have changed substantially since Elman introduced 'Elman nets' in 1990. These changes concern both the input that is used and the strategies that modelers use to train and test models. During the last two decades computational modelers also have developed a close co-operation with researchers working on child-language acquisition. In some cases (e.g. Christiansen, MacWhinney) the same researchers

work on both problems. This is important because findings from modeling can inform research on children, and new insights and data from child-research will allow us to develop better models. Because modeling has changed in many ways it is important to aim philosophical criticisms at clearly identified models. The 'one size fits all' approach advocated by McGilvray (2009) and Chomsky (2009) is not justified.

As the results of Blanchard show, a model that successfully simulates one aspect of language acquisition in one language may perform poorly on the same aspect when exposed to a different language. For this reason models should be tested in different languages, and/or findings from work on children in different linguistic communities need to be incorporated when designing models. Even though it is probably true that any child can learn any human language (Chomsky, 2005, 2007, McGilvray 2005, Smith 1999), it cannot be assumed that children acquiring different languages rely on the same set of cues. In this case computational modeling has shown that the hypothesis that language learning follows essentially the same steps regardless of the particular language learned needs to be re-evaluated. Some of the surprising results have shown that we cannot rely on *a priori* assumptions about language learning. Instead we need to develop testable hypotheses and models and adjust these in light of the obtained results.

The previous sections have shown that computational models can simulate some important aspects of language acquisition. Especially, models simulating the word-segmentation task perform reasonably well. Word segmentation is undoubtedly very important, but it is only one of many steps in the language acquisition process and it does not seem to require sophisticated cognitive capacities. Hence, the question

arises whether computational models can also handle more cognitively demanding aspects of language (e.g., syntax and semantics). In the next section I will address this question.

5.3.3.3. Modeling of Complex Aspects of Syntax and Semantics

Following the assumptions that motivated *Cartesian Linguistics,* one might argue that the more sophisticated aspects of language acquisition and language use (e.g., acquisition of semantics, syntax, and linguistic creativity) will remain forever beyond the capability of computational modeling. A complete discussion of this question would require addressing the problem of artificial intelligence and is beyond the scope of this work. Instead, I will highlight now some recent work involving the acquisition of complex aspects of syntax and semantics. These aspects have been used to support arguments for innate language-specific mechanisms/knowledge (e.g., Chomsky, 1966, 1975a, 1980, 1986b, 2000a,b, 2002, 2009a, 2010b; Smith, 1999; McGilvray, 2005, 2009).

Recent work in computational modeling has demonstrated that the acquisition of complex syntactical regularities can be simulated using bigram/trigram models (Christiansen & Chater, 1999; Reali & Christiansen, 2005) or simple recurrent networks (Elman, 1990; Elman, 1993; Cleermans, 1993; Christiansen & Devlin, 1997; Allen & Seidenberg, 1999; Christiansen & Chater, 1999; Mikkulainen & Mayberry, 1999; Pacton et al., 2001; Reali & Christiansen, 2005). These are

important findings because they show a possible alternative to a domain specific LAD. It is beyond the scope of this work to give a comprehensive overview of computational modeling (for some reviews see Christiansen & Chater, 2001; Christiansen et al. 2005; MacWhinney, 2010; Edelman & Waterfall, 2007). Therefore, I will highlight only some of the results of this work below. I begin with two models that simulate idiosyncratic peculiarities of French spelling and English grammar respectively. In both cases it becomes obvious that performance that seems to indicate explicit knowledge of abstract rules can be achieved by mechanisms that rely only on statistical features of the input.

Computer simulations of certain aspects of language acquisition are most useful when they model closely the relevant behaviour of children. For this reason Pacton et al. (2001) tested first whether children from kindergarten to grade five use statistical cues to track orthographic language regularities. They found that kindergarten children were able to use statistical regularities in the input to judge the 'word-likeness' of nonsense letter strings. Consistent with the statistical information available in the input, children judged letter strings as less word-like when they began with a double vowel than with a double consonant. Further, children from grade 2 onward made this distinction even for consonants that are never doubled. Pacton et al. duplicated these results with connectionist networks (SRNs) that were trained using back propagation of error mechanisms. Like the children, the SRNs became sensitive to the frequency and placement of doubled letters. The SRNs lack a mechanism for rule-based abstraction but still successfully simulate the performance of children. This led Pacton et al. to conclude that domain specific mechanisms for rule-based

abstraction are unnecessary to account for this aspect of language performance.

Another aspect of language acquisition that seems to push the limits of computational modeling is the acquisition of complex verb-argument structure, such as the prepositional dative (PD) vs. direct object dative (DOD). Many English verbs (e.g., tell, give, throw, bring) can occur in both constructions. For example 'to give' occurs in "Stefan gives a book to Katrina" (PD) and in "Stefan gives Katrina a book" (DOD). However, some verbs (e.g. confess, take, say, send) occur only in PD. Thus, when encountering a new verb, children cannot reliably generalize from previous examples because they do not know in which of the two groups the new verb belongs. Yet, seemingly, children are able to use novel verbs correctly, which has lead several researchers (e.g., Pinker, 1989; Gordon, 1990; Goldberg, 1985; Smith, 1999; McGilvray 2005) to conclude that some form of innate knowledge is required to explain this observed performance. Thus, the question arises whether computational models can achieve a child-like performance when acquiring PD and DOD constructions.

Perfors et al. (2010) present a domain-general hierarchical Bayesian model for the acquisition of PD and DOD construction. These researchers take as point of departure the work of Wonnacott et al. (2008). Wonnacott et al. found that human listeners are sensitive to distributional cues in language input and can use these cues to make productive generalizations. The computational model of Perfors et al. was informed by the knowledge that had been gathered from work with human children. They developed a computational model that explains "the acquisition of verb constructions as a rational statistical inference" (Perfors et al., 2010, p. 609).

Essentially this model can take advantage of positive and implicit negative evidence that is provided in the input. It keeps track of whether or not a given verb occurs in PD and DOD constructions. Assuming that the permissible usage is fixed, the model can use new data to make increasingly better predictions. "Each time a verb is encountered in one of the two competing structures, it is *not* encountered in the other, and this provides cumulative evidence against a grammar that allows this usage" (p. 630, original emphasis). This results in a learning outcome closely resembling that of human children: performance is very good for frequently occurring verbs but poor for verbs that are rarely encountered. Even though the model is capable of learning the distinction between alternating and non-alternating verb classes on the basis of syntactic input alone, the authors do not suggest that children exclude semantic information when learning this distinction. But the fact that it is possible for a relatively simple model to simulate this aspect of language acquisition suggests that the input contains an abundance of statistical information that can be used for inferences even before the learner has access to semantic information.

As we have seen, the modeling of some aspects of language acquisition has produced impressive results. Now I will focus on some models that have tackled more complex grammatical structures. This is an important area of computational modeling because it has been argued that complex grammatical structures could not have been learned without the help of an innate LAD because children know how to interpret these structures without training and relevant evidence (Chomsky, 1957, 1965, 1985; Crain & Nakayama, 1987; Crain 1991; Lightfoot, 1991; Smith, 1999; Crain & Pietroski, 2001; Crain & Pietroski 2002; Legate & Yang 2002; Baker 2005; Crain et

al., 2005). As we will see, the computational models that simulate aspects of syntax acquisition rely on mechanisms that are similar or virtually identical to those used in the initial steps of language learning (e.g., word-segmentation, acquisition of word categories, past tense acquisition). This could indicate that children and models rely on very similar resources in acquiring linguistic knowledge of increasing complexity. This does not prove that these resources are domain general. But it makes it less plausible to argue that they are *necessarily* domain specific.

Reali and Christiansen (2005) suggest that the language input contains rich indirect statistical information that can be accessed by the language learner. Using the example of auxiliary (AUX) fronting in complex polar interrogatives they show that simple learning devices, such as neural networks, are capable of exploiting such statistical cues. This is an important finding because the problem of polar interrogatives has played an important role in language acquisition debates since the 1960s. In English declarative sentences are turned into questions by fronting the correct auxiliary. Chomsky (1965a) used the following example to illustrate the problem. The declarative sentence '*The man who is hungry is ordering dinner.*' is grammatically turned into a question in (1a) but ungrammatically in (1b).

(1a) *Is the man who is hungry ordering dinner?*
(1b) **Is the man who hungry is ordering dinner?*

Because this fronting is not based on a structure-independent rule that could be readily learned from the available input but on a structure-dependent rule (Chomsky, 1980a), it has been argued that the knowledge allowing children to

produce correct auxiliary questions must be innate (e.g., Chomsky 1965, 1980; Crain & Nakayama, 1987; Crain 1991; Lightfoot, 1991; Smith, 1999; Crain & Pietroski, 2001; Crain & Pietroski 2002; Legate & Yang 2002; McGilvray, 2005).

Reali and Christiansen (2005) show a possible alternative to this suggestion. They trained simple statistical models based on pairs (bigrams) and triples (trigrams) of words of child-directed speech. Then they tested the models on sentences that consisted of correct polar interrogatives (e.g., *Is the man who is hungry ordering dinner?*) and incorrect ones (e.g., *Is the man who hungry is ordering dinner?*) that had not been present in the training corpus (Reali & Christiansen, 2005, p. 1010). They found that the models classified correctly 96 out of 100 grammatical test sentences and concluded that these "results indicate that it is possible to distinguish between grammatical and ungrammatical AUX questions based on the indirect statistical information in a noisy child-directed speech corpus containing no explicit examples of such constructions" (Ibid., p. 1014). Furthermore, their models were also able to simulate the production of grammatical AUX questions. This performance is based on frequency patterns in the input. Sentence chunks that are frequently encountered create a bias towards grammatical question-production even in the absence of direct positive evidence. Assuming that the models do not have innate knowledge of grammar, it seems to follow that the statistical information that is explicitly and implicitly available in the input can be used to produce grammatical AUX questions.

Many other researchers have reported similar results (e.g., Redington et al., 1998; Ellefson & Christiansen, 2000; Elman & Lewis, 2001; Mintz, 2002; Perruchet

& Vinter, 2002; Christiansen & Kirby, 2003). Again, it is important to establish that modeling closely mirrors abilities and strategies used by human learners. Mintz (2002) showed that adults who learned an artificial language naturally formed abstract grammatical categories solely based on distributional patterns of the input data. This could be evidence for rapidly engaged distributional mechanisms that also play a role in the early stages of language acquisition when the learner lacks access to other information (semantic and syntax). Mintz claims that his experiment "shows evidence of categorization mechanisms that function from distributional cues alone" (Mintz, 2002, p. 684). It has been shown that computational models can replicate other important aspects of human performance. Lewis and Elman (2001) trained simple recurrent networks on data from an artificial grammar. This generated questions of the form "AUX NP ADJ?" and sequences of the form "Ai NP Bi". During training the SRNs encountered no relevant examples of polar interrogatives. In this experiment it has been shown that the SRNs were better at making predictions for multi-clause questions involving correct auxiliary fronting than for those involving incorrect auxiliary fronting.

Christiansen and Kirby (2003) demonstrate that a general model of sequential learning that relies on the statistical properties of human languages can account for many aspects of language learning. Similar to the experiments on statistical learning discussed above, artificial-language-learning experiments showed that human subjects and SRNs that were trained on ungrammatical artificial languages made significantly more errors when predicting the next word of a string than subjects and SRNs that were trained on grammatical artificial languages. Languages are

288

considered grammatical when they contain (at least one of the) universal properties of natural languages (e.g., branching direction, subjacency) and ungrammatical when they lack these properties. The authors suggest that the close performance similarities of human subjects and SRNs could indicate that both rely on similar learning mechanisms.

Ellefson and Christiansen (2000) demonstrated that SRNs were significantly better at predicting the correct sequence of elements in a string of a 'natural language' than of an 'unnatural language'. The 'natural language' contained subjacency constraints, and the 'unnatural language' lacked these constraints. For example, SRNs trained on languages containing the 'natural' patterns exemplified in sentences (2) and (5) did significantly better than those trained on languages allowing the 'unnatural' patterns exemplified in (3) and (6).

(1) Sara heard (the) news that everybody likes cats.

(2) What (did) Sara hear that everybody likes?

(3) *What (did) Sara hear (the) news that everybody likes?

(4) Sara asked why everyone likes cats.

(5) Who (did) Sara ask why everyone likes cats?

(6) *What (did) Sara ask why everyone likes? (Ellefson & Christiansen, 2000, p. 349f).

Ellefson and Christiansen were able to show that SRNs trained on the same input data are sensitive to the statistical properties of the input and perform

significantly better on grammatical than on ungrammatical test sentences. Other researchers obtained similar results. For example, Perruchet and Vinter (2002) defend a plausible model (PARSER) of the acquisition of certain aspects of language based on the information that is contained in small chunks of the language input. They demonstrate that complex material can be processed as a succession of chunks that are comprised of a small number of primitives. According to these authors, associative learning mechanisms can fully account for this aspect of language learning.

When SRNs and other computational models are able to acquire statistical knowledge of the input based on positive examples alone, then it seems to be at least imaginable that children can pick up this information as well. Whether or not children rely on the same mechanisms as SRNs remains a point of debate (for some critical suggestions see Marcus, 1999; Marcus & Brent, 2003). But the existence of these mechanisms, again, casts some doubt on the *necessity* of an LAD.

We have seen that not only word-segmentation but also the acquisition of complex grammatical forms could be based on statistical properties of the input. Statistical learning occurs in several species. It has been confirmed in non-human animals (Ramas et al., 2000; Hauser et al., 2001; Terrace, 2001), and there is evidence that it might have been recruited to support language learning in human infants (Saffran et al., 1996; Fiser & Aslin, 2002; Maye et al., 2002). Whether or not statistical learning mechanisms can account for all aspects of language acquisition is a matter of ongoing debate. While this debate is far from over, it has become clear that proponents of the LAD need to rule out the possibility that data-driven general-

purpose learning mechanisms such as statistical learning can account for the acquisition of human language.

5.3.4. Recursion

In the final section of this chapter I will discuss some attempts to model one of the often-cited 'hallmarks' of human language: recursion. Recursion has played a central role in Chomsky's arguments for the uniqueness of language (Chomsky, 1957, 1966, 1975, 1980, 1986; Hauser, Chomsky, & Fitch, 2002; Fitch, Chomsky & Hauser, 2005). It allows for unbounded linguistic creativity and remains at the core of the Minimalist Program (Chomsky, 1995; Boeckx, 2006). Thus, the possibility that this unique feature of language can be simulated by computational models casts some doubt on the proposal that recursion is necessarily an "innate property of grammar or an a priori computational property of the neural systems subserving language" (Christiansen & MacDonald, 2009, p.127).

There are different types of recursive constructions as well as several levels of complexity within these types. Left- and right-branching recursion (LBR, RBR) are fairly common in many languages. One fairly complex examples of RBR in English is: "This is the dog, that chased the cat, that killed the rat, that ate the cheese, that lay in the house that Jack built" (Sampson, 2004, p.133). Even though this sentence involves 4 levels of RBR it can be processed and understood by an average native speaker and will generally be judged as grammatical.

Center embedded recursion (CER) on the other hand is more difficult to

process, as this example shows: "When the pain, that nobody, who has not experienced it can imagine, finally arrives, they can be taken aback by its severity" (Ibid., p. 20). This sentence only involves 2 levels of CER, yet it is difficult to process, and it takes a special effort to understand it. These differences between RBR/LBR and CER led numerous linguists to the belief that constructions containing higher-level CER are absent in human languages (e.g., Reich & Dell, 1977; Reich, 1969; Labov, 1973). While empirical research has shown by now that higher level CER constructions do occur in written and spoken language (for discussion see Sampson, 2001), it is generally accepted that these constructions are rarer than and judged as less grammatical than LBR/RBR at the same level of complexity. This raises an interesting problem for nativism. If recursion is at the heart of linguistic creativity and if language depends on an innate, genetically specified mechanism, then it is curious that different types of recursion pose distinctly different demands on language processing. Should we not expect that such closely related properties of language as LBR/RBR and CER are underwritten by very similar genetically specified mechanisms? This seemed indeed to be the default assumption of many nativists (e.g., Miller & Chomsky, 1963; Marcus, 1980; Church, 1982; Stabler, 1994; Chomsky, 2009b). It was proposed that the problems with multiple CER arise not from linguistic but from psychological mechanisms (e.g., memory and attention span limitations, difficulties to paraphrase and fluently read sentences with multiple CER, for discussion see Sampson, 2001; Christiansen & MacDonald, 2009).

Recently these assumptions have been challenged (e.g., MacWhinney, 2004; Christiansen, 1994; Sampson, 2001; Christiansen & Mac Donald, 2009). For these

reasons recursion modeling has two distinct purposes. First, if it can be shown that non-domain specific models can imitate the performance regarding recursive abilities of human speakers, the assumption that a domain-specific innate faculty is required for recursion is challenged. Second, if it can be shown that models that are not limited by memory and other non-linguistic factors process and comprehend LBR/RBR and CER in a similar way to humans, then it appears plausible to suggest that the differences are of a linguistic nature.

Chomskyans have often attacked computational models of language acquisition, and, maybe not surprisingly, models of recursion receive an especially hostile treatment as the following text illustrates:

> No matter how much computer power and statistics [connectionists] throw at the task, it always comes out ...wrong. Take Elman's famous paper ... on learning nested dependencies. Two problems: (1) the method works just as well on crossing dependencies, so doesn't bear on why language near universally has nested but not crossing dependencies. (2) His program works up to depth two, but fails totally on depth three. So it's about as interesting as a theory of arithmetical knowledge that handles the ability to add 2+2 but has to be completely revised for 2+3...Such approaches could do far better trying to duplicate bee communication or for that matter what's happening outside the window...Why don't they do it? Because it would be ludicrous. (Chomsky, quoted by McGilvray, 2009, p. 23).

The allegations made by Chomsky here pertain not just to one paper by one researcher. What is referred to as "Elman's famous paper" turned out to be a fictitious conglomerate based on the work discussed in several papers and books by several researchers (for details see Behme, 2009). This is important because Chomsky (2000,

293

2009) and McGilvray (2009) hold that the allegations do not pertain to the work of one individual researcher but to 'connectionists' in general. For this reason we need to look at more than one "famous paper" to evaluate whether or not connectionist models are 'ludicrous'. We will focus on a few of the many papers that have been published on this issue since the early 1990s.

Elman (1993) tested the hypothesis that "connectionist networks possess the requisite computational properties for modeling those aspects of natural language which are beyond the processing capacity of finite state automata [e.g. recursion]" (Elman, 1993, p. 75). From an artificial grammar he generated an input corpus of sentences with the following properties: (i) subject nouns and their verbs agreed for number; (b) verbs either required direct objects or optionally permitted direct objects or precluded direct objects; and (c) sentences could contain multiple embeddings in the form of relative clauses and subordinate clause (for full details see Elman, 1993, pp. 75 - 77). The network was trained to take one word at a time and predict what the next word would be. To make correct predictions the network needs to represent internally grammatical dependencies of the input. Elman found that networks that were trained on input of slowly increasing complexity achieved high performance and "generalized to a variety of novel sentences which systematically test the capacity to predict grammatically correct forms across a range of different structures" (Ibid., p. 77). On the other hand, networks that had access to the entire input corpus at once performed very poorly. Finally, networks that had been exposed to the complete input corpus from the beginning but were given a slowly increasing memory capacity (for details see pp. 78-79) had a prolonged initial learning phase but performed very well

after that.

Elman claims that children will probably neither encounter the first nor the second condition. That is the input is never carefully matched to the language acquisition stage the child is at (condition 1), and the child is never able to make use of all the information that is contained in the complete input (condition 2). But it is plausible to assume that when children are learning, they are in a similar position to that of the networks in condition 3: they are faced with complex input, but their ability to access the input is limited. Over time this ability improves just as in the networks where "the learning mechanism itself was allowed to undergo "maturational changes" (in this case, increasing its memory capacity) during learning" (Ibid., p. 79).

There are two important points here. First, it could be shown that networks are able to 'learn' even such complex grammatical structures as CER from the input. This, again, indicates that a domain-specific mechanism may not be required to achieve this result. At the very least it shows that the statistical information that is present in the input could be sufficient for the acquisition of CERs. Second, the fact that networks that are 'handicapped' in some way (by exposure to limited input or limited memory capacity) are more successful 'learners' than networks that have access to the complete input corpus and maximal memory capacity from the beginning had not been predicted by the experimenter. It was actually necessary to perform the experiments to obtain "a deeper understanding of the principles which constrain learning in networks" (Ibid., p. 85). As our understanding of these principles improves, our ability to develop better computational models and relevant test procedures for children improves as well.

Regarding the specific criticism of Chomsky/McGilvray, it would seem that there is no evidence that Elman's model "works up to depth two, but fails totally on depth three" (Chomsky quoted in McGilvray 2009, p. 23). The models discussed by Elman either achieved poor performance across the spectrum of simple and complex sentences, or they achieved good performance in both. Models that performed poorly performed so across the board, but Elman did not report a *total* failure specifically at depth three.

The concern that "the method works just as well on crossing dependencies, so doesn't bear on why language near universally has nested but not crossing dependencies" (Ibid.) has not been addressed by the Elman's (1993) work because Elman did not compare how models performed on crossing vs. nested dependencies. Other researchers have attempted to do this comparative work, and to their models we shall turn next.

Christiansen (1994) trained SRNs on a recursive artificial language and found that performance differed for higher levels of complexity of RBR and CER. Christiansen and Chater (1999) attempted to model the human processing performance for RBR, CER and cross-dependency recursion (CDR). For this purpose they trained "connectionist networks on small artificial languages, which exhibit the different types of recursive structure found in natural language" (Christiansen & Chater, 1999, p. 159). They found that the networks performed well on RBR and single-embedded CER and CDR but that performance quickly degraded for CER and CDR when additional levels of embedding were added. Christiansen and Chater suggest that what "constrains the performance of the SRN appears to be architectural

limitations interacting with the statistics of the recursive structures" (Ibid., p. 172).

They also observed that, even though the models were not trained on constructions of

recursive depth four, "there was no abrupt breakdown in performance for any of the

three languages at this point... This suggests that these models are able to generalize

to at least one extra level of recursion beyond what they have been exposed to during

training" (Ibid., p. 182).

The performance of SRNs closely resembles that of humans, who also can

process several levels of RBR but have difficulties processing doubly or more highly

embedded CER and CDR. Christiansen and Chater (1999) observe that the difficulty

of processing center-embedded structures is not confined to a linguistic context. They

cite Larkins & Burns (1977), who demonstrated that when subjects were asked to

name center-embedded pairs of letters and digits, they experienced the same difficulty

as when processing center-embedded sentences. Christiansen and Chater hypothesize

that non-linguistic processing constraints could be to blame for the poor performance

in the CER/CDR tasks. The findings that SRNs, which are not language specific,

show the same performance limitations could support this hypothesis. These results

also have consequences for the nativist/empiricist debate. Christiansen and Chater

propose that

> These results suggest a reevaluation of Chomsky's (1957, 1959)
> arguments that the existence of recursive structures in language rules out
> finite state and associative models of language processing. These
> arguments have been taken to indicate that connectionist networks, which
> learn according to associative principles, cannot in principle account for
> human language processing. But we have shown that this in principle
> argument is not correct: Connectionist networks can learn to handle
> recursion with a comparable level of performance to the human language

processor. (Christiansen & Chater, 1999, p.199).

I would agree that the recent work in modeling has provided some reason for questioning the cogency of 'in principle' arguments against connectionism. It is also worth noting that the arguments against connectionist modeling used by Chomsky and McGilvray in 2009 have not changed in character from those used by Chomsky in 1957. Either engagement with actual work is completely avoided, or, when such work is cited ('Elman's famous paper'), it is misrepresented. It appears implausible that the misrepresentation is based on genuine misunderstanding of the work discussed. For example the challenge that connectionist models are not relevant to human languages because "the method works just as well on crossing dependencies" (McGilvray, 2009, p.23) has been shown incorrect in all cases we have evaluated here. Equally, the claim that the "program works up to depth two, but fails totally on depth three" (Ibid) is incorrect. What Christiansen and Chater (1998) reported is that the performance of the models *declines* for higher-level embeddings, not that it fails totally. In fact, Christiansen & Chater's model did not even fail completely at depth four. Given that these results have been available for a decade, it is somewhat surprising that Chomsky and McGilvray still repeat the 'in principle' arguments of the 1950s.

It might be argued that, nevertheless, some aspects of the models introduced so far are problematic. For example, the training procedures do not reflect the language-acquisition process in children, and the input may contain more examples of RBR, CER and CDR structures than input to which children are exposed.

Christiansen and Chater do not claim that their model replicates the acquisition process in children. Their goal was to show that the statistical information contained in the input is sufficient for *some* mechanism to acquire the ability to process recursive structures. This goal has been achieved.

Other work has also shown that the language input contains sufficient implicit information to allow for the acquisition of recursive structures. Christiansen and MacDonald (2009) show that a connectionist model is capable of simulating human performance of processing complex center-embeddings in German and cross-dependencies in Dutch. As previously discussed, these recursive constructions are more difficult to process than the simpler, right- and left-recursive structures. The authors suggest, "the ability to process recursive structure is acquired gradually, in an item-based fashion given experience with specific recursive constructions" (Christiansen & MacDonald, 2009, p.127). The SRN was trained on an artificial context-free grammar with a 38-word vocabulary (for details see Ibid., pp. 130-132) and then tested on novel sentences. The SRN was able to acquire complex grammatical regularities and "to make nonlocal generalizations based on the structural regularities in the training corpus" (Ibid., p. 132). Another important finding was that for both humans and SRNs doubly embedded CER were harder to process than doubly embedded CDR. This confirms the finding of Christiansen & Chater, 1999). The close fit to human data also extended to novel predictions where the models made grammaticality judgments similar to those of human subjects (Christiansen & MacDonald, 2009, p.149). Obviously there are some limitations to these models. The vocabulary is very small, and only a small subset of grammatical

regularities is covered. Whether or not future models can be scaled up to the full complexity of human language remains to be seen.

Models that simulate the conditions under which human children learn language perform similarly to children. When children acquire language, the processing and production of recursive structure emerges gradually over time (Dickinson, 1987). This is contrary to Chomsky's assumption (1980) that the processing and production of recursive structure is virtually instantaneous. The work of Christiansen and MacDonald (2009) showed that, just like human children, SRNs do have a learning curve and their performance improves over time. Overall computational modeling of the acquisition of recursive structure has shown that there are several parallels between the performance of children and SRNs. This fact alone, of course, does not prove that the mechanisms exploited by SRNs are the same as those used by children. But the findings discussed here indicate that an innate domain-specific mechanism is not necessary to account for the acquisition of these structures. And the results obtained can direct future research and, it is hoped, provide insights into the mechanisms that allow human children to acquire language.

5. 4. Conclusions

In this chapter I have discussed computational models of language acquisition and evaluated whether or not the criticisms proposed by Chomsky (2000, 2009) and McGilvray (2009) apply. Based on the models I discussed, I suggest that the main points of criticism are not justified. Below I briefly summarize my findings.

McGilvray alleges that connectionists hold that neural nets (and by extension the models they use to simulate those nets) are initially "undifferentiated, approximating Locke's blank slate" (McGilvray, 2009, p. 110). I have found no evidence supporting this allegation. All researchers I discussed agree that there is some structure initially built into their models and that this structure is relevant to the performance of the models. What is at issue for them is not whether or not neural nets are initially structured (they are) but whether or not this structure supports domain-specific learning or learning across different domains.

McGilvray holds that connectionist modeling is irrelevant to language learning in humans. He suggests that the models rely on methods and mechanisms that are completely different from those of language learning: "the empiricists must assume that much of the child's early life ... is devoted to focused data gathering and training sessions" (McGilvray, 2009, p. 16). It seems that McGilvray is conflating several issues here. First he still relies on the Chomskyan dictum that language acquisition in children is effortless ("effortless creative use of language on the part of all normal three-and-a-half or four-year-old children" (Ibid., p. 17)) and virtually instantaneous. We have seen in chapter 4 that these claims are not supported by empirical research.

Second, he may seriously overestimate the time it takes the models to 'learn' something. Presumably he still relies on the observation of a fellow Chomskyan that "it takes a quarter of a million repetitions to train [a connectionist net] to recognize *Boys like girls* as well-formed" (Smith, 1999, p. 132). Like McGilvray, Smith does not reveal his source, so all we know is that he must refer to work that has been

completed before 1999. Assuming that Smith's claim was correct for the model he refers to, are we justified in further assuming that it can be extended to all work that has been completed until the time of McGilvray's publication in 2009? I think we are not. Current models vary widely in their demands for data, and it would be up to critics to cite specific models that, in their opinion, rely on implausibly large data input. The models I have discussed require less data than children who learn language. For example, Christiansen and Mac Donald (2009) state that "the SRN was trained via a word-by-word prediction task on 50,000 sentences" (Christiansen & Mac Donald, 2009, p. 131), Monaghan and Christiansen (2010) used up to 10,000 utterances from corpora containing 8313 to 27,794 utterances (Monaghan & Christiansen, 2010, p. 553) to train their PUDDLE model, Ryttig et al. (2010) used for their simulation training corpora containing between 6443 and 22,198 words (2861 to 7030 utterances, Ryttig et al., 2010, p. 527). All these corpora are substantially smaller than the input children receive when facing similar tasks to those faced by the models.

McGilvray claims that empiricist work involves a "massive amount of question begging ... [and that] Chomsky has noted many of these gaps and others in empiricist views since at least 1957..." (McGilvray, 2009, p. 17). I do not have the space here to go through his list of rhetorical questions one by one but want to address a few.

3.a. "How does the child manage to develop the notion of a phoneme?" As empiricist work has shown, it is not necessary to rely on an explicit notion of phoneme to succeed in the word segmentation tasks. The statistical information

present in the input is sufficient for the success of segmentation models. Thus, it would be up to McGilvray to show that children actually *do* rely on this notion and not on easily accessible statistical regularities when they acquire language.

3b. "How specifically does analogy work in trying to extend acquired knowledge to unacquired?" Computational models show that it is possible to apply what has been learned from the training set to the test set without the need of explicit knowledge of the underlying structural rules. It would seem at least possible that children learn in a similar way, and McGilvray needs to show that they do not.

3.c. "What counts as sufficient to show that child N has acquired any or all concepts $\{c_1...c_n\}$ or rules $\{R_1...R_n\}$?" Again the computational work shows that the models can perform *as if* they had learned rules when clearly they did not. It would seem that McGilvray fails to draw the distinction between rule-following and rule-obeying behaviour.

In some cases modeling had results that did not conform to the predictions made prior to modeling. In these cases the obtained results influenced future research. First the experiments were repeated. When the repetition confirmed the earlier findings the theoretical assumption on which the models originally had been based were adjusted. This is contrary to McGilvray's accusations "that dogma, not reason, drives the empiricist research strategy" (McGilvray, 2009, p. 23) and that 'empiricists' efforts like these make no contribution to sciences of the mind" (Ibid., p. 24). Asking whether or not an individual empiricist model makes a worthwhile contribution to science is of course a legitimate question. But the answer can only be found by evaluating the model, not by *a priori* assumptions about 'empiricist

models'.

In spite of some impressive successes, computational models are still a considerable distance away from simulating all aspects of language acquisition. Recently work has begun to simulate more complex aspects of language acquisition from multiple-cue integration (e.g., Christiansen, et al., 2010; Rytting et al., 2010; Christiansen & MacDonald, 2009; Onnis et al., 2009; Monaghan & Christiansen, 2008), and language acquisition in different languages (e.g., Freudenthal et al., 2010; Jaroz, 2010; Blanchard et al., 2010; Christiansen & MacDonald, 2009), word-sense disambiguation (e.g., Waterfall et al., 2007) to the construction of a complete, empirical, generative model of the learning of syntactic patterns (Waterfall et al., 2010). Yet, we are still a considerable distance away from any model that simulates the complete process of language acquisition. Thus, it remains an open question whether or not it will be eventually possible to combine many of the current 'small scale' models or if the complexity of the task exceeds the ability of data-driven models. Nevertheless, the work completed so far has provided valuable insights, and I disagree with McGilvray's evaluation that "empiricist efforts like these make no contribution to sciences of the mind" (p.24).

It has become evident that the remaining open questions cannot be answered by *a priori* theorizing, and it remains important that the models "can be tested, refined or rejected on the basis of publicly available data and/or replicable experiments" (MacWhinney, 2010, p. 14). And the data used for evaluating models need to be those collected from the interactions of real children and their language communities.

Chapter 6. Conclusions

In this thesis I have evaluated some of the contributions of Noam Chomsky's work to linguistics. Using the framework of his *Cartesian Linguistics* I have inquired weather he has placed his work in the context of a rich tradition of rationalist Cartesian thinkers, whether his work has contributed to clarifying important linguistics concepts and has placed linguistics within the natural sciences and whether his harsh criticism of empiricist linguistics and computational modeling of language acquisition is justified. Below I will summarize my findings and offer some suggestions for future research.

6.1. The Cartesian Tradition

Chomsky motivated the revival of alleged Cartesian roots of his own theorizing by claiming that the contributions of earlier linguists could provide insights relevant to contemporary problems. He suggested: "this return to classical concerns has led to a rediscovery of much that was well understood in…the period of Cartesian Linguistics" (Chomsky, 2009, p. 57). Chomsky acknowledges that "Descartes himself devoted little attention to language and his few remarks are subject to various interpretation" (p. 58). Yet he justifies the label 'Cartesian' by claiming that during the Cartesian period there was "a coherent and fruitful development of a body of ideas as an outgrowth of the Cartesian revolution" (p. 58).

The claim that there existed such a coherent body of ideas related to the

Cartesian revolution has been challenged on historic grounds (e.g. Miel, 1969; Lakoff, 1969; Aarsleff, 1970, 1971; Percival, 1972; Gipper&Schmitter, 1979). Chomsky has never refuted these challenges, but has held instead that his methodology as a historian is preferable over more traditional approaches. First, Chomsky advocates what could be called the 'selective-history-approach' (SHA): "One might say that I'm looking at history ... from the point of view of … an art lover who wants to look at the 17th century to find in it things that are of particular value and that obtain part of their value … because of the perspective with which he approaches them" (Chomsky, 1971). Later Chomsky adds what I would call the 'rewrite-history-approach' (RHA): "The first [question], the actual sequence of events, is not in itself very interesting in my opinion; it's a story of chance events and personal accidents, accidents of personal history. The second question, namely, *how it should have happened*, is far more interesting and important, and that certainly has never been told or even investigated" (Chomsky, 1997, emphasis added). Combined, SHA and RHA allow Chomsky to pick and choose what he considers of value in Descartes' (and other rationalist/romantic predecessors') writings and to transform other passages into what Descartes/Rationalists should have written. This might seem to justify the artistic freedom Chomsky applies to history. However, it also shows that Chomsky has little interest in the facts of history, but intends to use the suitably re-interpreted Cartesians as figurants or ventriloquist puppets on the Chomskyan-Linguistics stage. For this reason I suggest that it is misleading to call Chomsky's work *Cartesian* Linguistics.

If we *are* interested in the historic roots of current linguistic thought, we

ought to pay close attention to the contribution of writers like Descartes to linguistic theorizing. This should be in Chomsky's spirit as he specifically claimed that one important task for *Cartesian Linguistics* "is to determine the exact nature of the 'capital of ideas' accumulated in the premodern period…and to find ways to exploit it for advancing the study of language" (p. 58). In philosophy the tight connection of current work to earlier theorizing has long been acknowledged, "…philosophy does not outgrow its dependence on the thought of the great figures of the past" (Katz, 1986, p. 6). Thus, it seemed appropriate to evaluate this mostly neglected capital of Cartesian ideas, beginning with the ideas of Descartes.

Two points seemed especially of importance. First would Descartes have held views that reasonably could be taken as foreshadowing the main commitments of Chomskyan linguistics? And, second, would Descartes have approved of the 'polemical elements' that "elevate 'Cartesian' approaches to the study of language and ...depreciate [all other approaches]" (Aarsleff, 1971, p. 570)? My literature research has shown that the answer to both questions is 'no'.

Specifically, I have shown that it is dubious that Chomsky's linguistic work can be traced back to a coherent rationalist tradition of which Descartes was one important founder. On the one hand we find several statements from Descartes that appear to support a very different judgment regarding his views of language acquisition. On the other hand several 'empiricist philosophers' held views that were more similar to Chomsky's view than the views of Descartes. I provide examples for both in turn:

When for example on hearing that the word "K-I-N-G" signifies

307

supreme power, I commit this to my memory and then subsequently recall the meaning by means of my memory, it must be intellectual memory that makes this possible For there is no relationship between the four letters (K-I-N-G), which would enable me to derive the meaning from the letters. It is intellectual memory that enables me to recall what the letters stand for (CSM III, pp. 336-7)

Here Descartes' emphasis clearly is on associationist learning of the sort usually associated with empiricism, not on innate knowledge. Even when we look closely at some of the passages cited by Chomsky, it becomes evident that Descartes focuses not only on creativity of language use, but also on behaviorist criteria for testing whether or not an organism is intelligent. One such example is: "men born deaf and dumb...usually invent their own signs to make themselves understood by those who being regularly in their company have the time to learn their language" (CSMK, III, p. 303; cited without reference in Chomsky, 2009, p. 60). Here the emphasis seems to be on communication, because the success criterion is "being understood". I have discussed several similar examples in chapter 2 and suggest that a truly Cartesian account of language acquisition would have to include Descartes' 'empiricist' commitments: language is used for communication and learned at least in part by associationist strategies.

Similarly, I found no indication that Descartes would support a domain-specific language faculty. For him minds are indivisible: "we cannot understand a mind except as being indivisible. For we cannot conceive of half a mind" (CSM II, p. 9). Furthermore, Descartes states that our knowledge depends only on a purely spiritual power which is "one single power...it is one and the same power. ...According to its different functions... the same power is called either pure intellect

or imagination or memory or sense perception" (CSM I, p.42, emphasis added). It has become evident that the belief of the actual Descartes, that the mind is indivisible, was not based on 'irrational dogmatism' (Chomsky, 1980, p. 30). This belief was justifiable in his overall philosophical framework and should be acknowledged as his legitimate belief. There is evidence from Descartes' own writing that he would not have held on dogmatically to his beliefs: "I would be sinning against good sense if I were to take my previous approval of something as obliging me to regard it as good later on, when it had perhaps ceased to be good" (CSM I, p. 123). Thus we have some reason to believe that, in light of new evidence, Descartes could have changed even his most fundamental beliefs (e.g., that a non-deceiving God necessarily exists, mind-body dualism, that animals do not have minds). But *how* he would have changed these beliefs is a matter of speculation. When we wish to engage in such speculation, we should clearly say that this is what we are doing, and refrain from ascribing to Descartes beliefs that he clearly did not hold.

Critics and supporters of Chomsky alike seem to accept that Descartes' actual contributions to linguistics are negligible. This is understandable; one finds nothing resembling a comprehensive linguistic theory in Descartes' writing. Nevertheless, many interpreters (e.g., Chomsky, 1966; Barsky, 1997; Cowie, 1999; Matthews, 2005, McGilvray, 2005; 2009) are confident that Descartes' doctrine of innate ideas would support a linguistic nativism that is similar to Chomsky's. My reading of Descartes suggests that his view of innate ideas would not have committed him to a Chomskyan theory of language acquisition. Specifically

Cartesian innate ideas do not seem to underwrite the effortless learning of semantic concepts and grammatical rules that are the hallmark of Chomsky's nativism.

Chomsky claims Descartes' commitment to innate ideas is similar in spirit to his own. But it is evident that he pays very little attention to what Descartes wrote specifically on the acquisition of language and the difference between ideas we acquire in our childhood and later in life. In a letter to Mersenne Descartes asserts that "[t]here are only two things *to learn* in any language: the meaning of words and grammar" (CSMK, III, p. 10, emphasis added) and in a letter to Chanut he explains that when we acquire the meaning of words we establish an essential connection between sense perception (objects and sound combinations) and mental concepts. Concerning the content of ideas Descartes repeatedly points out that the ideas we acquire in our childhood are often "confused thoughts" (CSMK, III, p. 308), based on "false preconceptions" (CSMK, III, p. 233). Usually we learn only much later about the true nature of things and acquire what Descartes calls 'clear and distinct ideas'.

One illuminating example for the difference between obscure and clear and distinct ideas can be found in the discussion of our different ideas of the sun. One of these ideas, depicting the sun as small, is derived from the senses and the other, depicting the sun as several larger than the earth, "is derived from certain notions that are innate in me" (CSM, II, p. 27). Relevant to language acquisition is the fact that children probably have only the idea that is derived from the senses. Of course, that does not prevent them from using the word 'sun' in appropriate circumstances, creatively and stimulus free. Even young children understand that the word 'sun'

refers to a certain object, and they do not need to see the sun every time they talk about it. Descartes' attempts to distinguish true science from everyday (mis)conceptions indicate that he is quite aware of the fact that our language use does not rely on a deep understanding of the true nature of things.

A similar sentiment is expressed in *Search for Truth* when Descartes refers to learning in early childhood: "...he came into the world in ignorance and ... the knowledge he had as a child was based *solely* on the weak foundation of the senses and the authority of his teachers" (CSM II, p. 400, emphasis added). Part of the knowledge acquired in early childhood is undoubtedly knowledge of language. And while Descartes makes it clear that science could not be placed on the shaky foundation of empirical knowledge he does not seem to have similar concerns about language.

Descartes seemingly allows also an 'empirical' account of concept acquisition: "...when we *see* a figure made up of three lines we *form an idea* of it which we call the idea of a triangle; and we make later use of it as a universal idea..." (CSM I, p. 212, emphasis added). According to this passage seeing an object (triangle) is sufficient not only for remembering that object later but also for forming a universal idea which we can apply later to objects of the same kind (other triangles). Similarly, when referring to numbers, Descartes claims that when considered in the abstract, "number ... is nothing more than a mode of thinking; and the same applies to all the other things which we call 'universals'" (CSM I, 58; AT VIIIA, 27). Further, "universals arise solely from the fact that we use one and the same idea for thinking of all individual things which are similar to each other..."

(CSM I, 59; AT VIIIA, 27).

Based on this evidence it seems reasonable to assume that for Descartes language learning could depend to a large degree on sense perception, association, and memory. He states that language "can be acquired without any process of reasoning... [based] on experience alone" (CSM, II, p. 403), that we learn language by connecting words with their meanings and remembering later, upon encountering words, which things they signify and vice versa (CSMK, III, p. 307) at a time when our thoughts are 'confused' and based on 'misconceptions', and that universals can be acquired by repeatedly observing particulars and deducing relevant similarities (CSM I, p. 212). However, it is also important to remember, that for Descartes only humans can learn language. For this reason language acquisition cannot be a purely mechanical process as Chomsky (2010) suggests. There are obvious references to interactions between the senses and the mind during language acquisition (e.g., CSM I, p. 220; CSM I, p. 348; CSMK III, p. 12), and Descartes repeatedly rejects the idea that a purely mechanical entity (automaton or animal) could acquire human language (e.g., CSM I, p. 139; CSMK III, p. 99).

When innate ideas play a role for language acquisition in Descartes' account, we notice another difference to Chomsky's view. For Chomsky our innate endowment guarantees that we acquire semantic concepts and grammatical knowledge effortlessly and virtually uniformly across the species. For Descartes innate ideas do not give similar guarantees. This is evident in his discussion of the idea of God. On the one hand Descartes asserts that we all have an innate idea of God. Yet, there are great differences in our ability to access this idea. Some people

(atheists) believe that "all we can understand about God is what he is called, namely 'God', or what corporeal forms painters use to represent him" (CSM I, p. 305). While Descartes believes that the atheists are utterly mistaken about the essence of God, he does not imply that they merely misunderstand the word 'God'. In fact, it is essential that the atheist understands (at least to some degree) what he denies.

In *Discourse on Method* Descartes observes that "...many are convinced that there is some difficulty in knowing God... The reason for this is that they never raise their minds above things which can be perceived by the senses" (CSM I, p. 129). Here it appears as if Descartes claims that sense experience is not only insufficient for understanding God but also utterly irrelevant to such an understanding. While this might commit Descartes to some extreme innatism regarding the content of the idea of God, it appears to be a very different type of innatism from Chomskyan innatism regarding language acquisition.

In *Principles of Philosophy* Descartes admits that, unless we focus our attention on the contemplation of God, we can be misled by the "habit of distinguishing essence from existence...[and doubt] whether the idea of God is not one of those which we made up at will, or at least one of those which do not include existence in their essence" (CSM I, p. 198). This indicates that having the innate idea of God does not guarantee having a correct understanding of the concept. However, Descartes never expresses any doubt regarding the ability of anyone to use the word 'God' appropriately. Clearly, the way in which the atheist or the inattentive believer uses the word 'God' is different (in kind) from the way in which the magpie uses the words 'good day'.

There is a further important difference between Descartes' and Chomsky's views regarding the contribution of innate ideas/knowledge to language acquisition. When Descartes speaks about innate ideas, he refers either to operations of the mind or to (semantic) content of ideas. He says virtually nothing about syntactic rules of language and how they are acquired. By contrast, for Chomsky what is innate are syntactical or structural principles that could not have been acquired from the available input. These structures are usually complex, not simple.

Finally, Descartes suggests that the fact that (the content of) an idea is innate in and of itself does not guarantee that we have explicit knowledge of it. This is a notion of 'innate' very different from Chomsky's: for Chomsky the innate language faculty allows children to learn language effortlessly, often even without being exposed to examples of what they learn. For Descartes it is possible to have an innate idea of God without becoming knowledgeable of it even after reading the '*Meditations*' a thousand times.

Thus, on one hand we have seen that for Descartes first language acquisition could rely mainly on the linguistic input perceived by the child, associationist learning and memory. On the other hand, if for Descartes innate ideas play a role in language acquisition at all, their role is very different from the role played by the innate endowment of Chomskyan linguistics. Finally, there is no textual evidence suggesting that Descartes would believe that language is underwritten by a domain-specific mechanism that is decoupled from 'general intelligence'. These facts suggest that it is misleading to call Chomsky's linguistics Cartesian.

Turning now to some empiricist predecessors of the Cartesian and post

Cartesian period, we can see that several elements of Chomskyan linguistics could be traced back to their work. For example, John Locke clearly emphasized the fundamental distinction between human and animal communication:

> Brutes abstract not. If it may be doubted whether beasts compound and enlarge their ideas that way to any degree; this, I think, I may be positive in,- that the power of abstracting is not at all in them; and that the having of general ideas is that which puts a perfect distinction betwixt man and brutes, and is an excellency which the faculties of brutes do by no means attain to. For it is evident we observe no footsteps in them of making use of general signs for universal ideas; from which we have reason to imagine that they have not the faculty of abstracting, or making general ideas, since they have no use of words, or any other general signs. (Locke, Book 2, Chapter XI, 10)

There were debates about the species-specificity of language in Cartesian times. But it is by no means clear that the opponents of the debates were empiricists vs. rationalists. From a philosophical point of view it might be an interesting task for future research to analyze the arguments for species-specificity used in Cartesian times and compare them to those used in current debates.

Regarding historic antecedents of Chomskyan ideas, in the writings of another empiricist we find a more succinct expression of the belief in language universals than anywhere in Descartes' writings:

> Among different languages, even where we suspect the least connexion or communication, it is found, that the words, expressive of ideas, the most compounded, do yet nearly correspond to each other: a certain proof that the simple ideas, comprehended in the compound ones were bound together by some universal principle, which had an equal influence on all mankind. (Hume, 1777/1958, pp. 22-3)

315

Examples like these indicate that there was no specifically and uniquely rationalist linguistic tradition foreshadowing the commitments associated with Chomsky's linguistics. They suggest that there was a much richer and more diverse tradition of linguistic thought in the Cartesian period and that the divide between rationalists and empiricists was by no means as hard and fast as suggested by Chomsky (1966) and McGilvray (2009). Thus, there are no clear historic antecedents for the separation of linguistic research methods introduced in *Cartesian Linguistics*: "Those who Chomsky thinks can plausibly deal with the issues that linguistic creativity poses for the mind he calls 'rationalists'; those who cannot, he calls 'empiricists"'" (McGilvray, 2009, p. 1). This perceived divide has permeated much of Chomsky's work, recent work in the biolinguitsic enterprise (Di Sciullo & Boeckx, in press) and in Cartesian biolinguistics (Boeckx, 2009, in press). However, it is anything but Cartesian in spirit.

Descartes held that "there are two ways of arriving at knowledge of things – through experience and through deduction" (CSM I, p. 12). Thus Cartesian science relied on both: sense experience and deduction. And while he emphasizes that 'pure inference' can never be performed wrongly by a rational mind, he also cautions that "the errors to which men…are liable …are due only to the fact that men take for granted certain poorly understood observations, or lay down rash and groundless judgments" (Ibid.). It would appear that insisting on an insurmountable divide between the empiricist and rationalist research programs is based on the kind of rash and groundless judgments Descartes cautions against.

Furthermore, Descartes warns that we should "not take refuge in ambiguity"

(CSM II, p. 273) but need to "communicate faithfully to the public what…[we] had discovered and … urge the best minds to try and make further progress by helping with the necessary observation" (CSM I, p. 143). As we have seen, Chomsky's work does not live up to these requirements, which are as essential for contemporary scientific collaboration as they were at Descartes' times. Regarding his own conduct Descartes wrote that he always needs to remember

> …that there might be some mistake in one of my own theories, in spite of the great care I had taken never to adopt any new opinion for which I had no certain demonstration, and never to write anything that might work to anyone's disadvantage (CSM I, p. 142)

Adopting these principles would be beneficial to the Chomskyan enterprise. Finally, Descartes writes: "I would be sinning against good sense if I were to take my previous approval of something as obliging me to regard it as good later on, when it had perhaps ceased to be good" (CSM I, p. 123). This can be interpreted as a warning against insisting on previously held principles, after it has been shown that these principles are problematic. Overall then, neither Chomsky's linguistics nor his research methodology can be traced back to Cartesian roots if we take 'Cartesian' to refer to something that would have been endorsed by Descartes.

6.2. *Chomskyan Progress*

According to McGilvray (2009), Chomsky's linguistic science has made

steady progress towards a better understanding of "the inner mental machinery that puts concepts in place or activates them, configures them in forms that the machinery allows or requires, and does the same for the rules …that govern how to put concepts together" (p. 6). In the 1950s Chomsky's promise to bring rigour and exactness to linguistics and to situate linguistic theorizing firmly within the natural sciences (Chomsky, 1957, 1965, 1966) clearly revived linguistics. However, 60 years later many of the initial promises remain unfulfilled. Chomskyan science remains vague about the mental machinery that underwrites creative language use and has not provided testable hypotheses regarding the mechanisms that allow for language acquisition.

Chomsky himself has claimed recently that an understanding of language use and creativity may lie beyond our cognitive capacities: "… the Cartesian question of creative [language] use…remains as much of a mystery now as it did centuries ago and may turn out to be one of those ultimate secrets that will ever remain in obscurity, impenetrable to human intelligence" (Chomsky, 2010, p. 29). This claim, which is not justified by reference to any research program, stands in stark contrast to McGilvray's celebration of steady progress of Chomskyan science. But, as my research has shown, it may be a more accurate evaluation of the tangible scientific progress that Chomsky has made. However, I did not find evidence supporting Chomsky's implied conclusion that, because he was unable to uncover the mechanisms that underwrite creative language use, this task is beyond human capacities.

Early in his linguistic career Chomsky focused on syntax and grammar (e.g.,

Chomsky 1956, 1957 1959, 1965, 1966, 1968). Some of the key claims of Chomsky's early work were (i) that human languages have syntactic universals, (ii) that these universals determine a well-defined class of grammatical sentences, and (iii) that it is possible to find a formal procedure (grammar) to generate this class of grammatical sentences. A constituent grammar provides a finite set of rules for combining constituents into larger units (sentences) and can generate an infinitely large language. Chomsky claimed that human languages cannot be generated by constituency grammars alone and proposed that an additional series of transformational rules is needed to generate all grammatical sentences of human languages. This early work contributed to clarifying important conceptual issues, provided a scientific framework for linguistics and had an impact that reached far beyond linguistics.

Within the next 20 years Chomsky continued to focus on syntax and refined his theories. He proposed that what we commonly perceive as language is in fact only a 'surface structure' that reflects indirectly an underlying deep structure; the true subject of linguistic inquiry (Chomsky, 1970, 1973). To explicate the relationship between deep and surface structure Chomsky conceived of transformational rules of ever increasing complexity that convert deep structures to surface structures. Other conceptual advances were the introduction of the competence and performance and the E- and I-language distinction. All these moves emphasized the fact that the object of linguistic study should be physical parts of biological brains (the putative language organs). Chomsky asserted that the focus of non-Chomskyan linguists on E-language is essentially fruitless. For him E-language

was an arbitrary, artificial construct "that is understood independently of the properties of the mind/brain" (Chomsky, 1986, p. 29) and their study would not reveal anything interesting about the nature of language.

If we conceive of language as some well-defined part of the physical brain, then the brain should be the object of linguistic study. Chomsky himself was famous for the comment "I hate experiments" (Mehta, 1971, p. 209), and it is usual practice in the well-established sciences that some people specialize in theorizing and others in experiment. However, it appears that Chomskyans are virtually all theorizers and no experimenters. Thus, it has not been established yet how the theories are tied to observable reality at all. Further, Chomsky's own work to this point has not contributed directly to locating language in the brain. He continues to work 'on the surface' of alleged epiphenomena and attempts to infer deep structure from analysis of samples of sentences. Thus, over decades Chomskyans and empiricists alike have refined their theories, but, unlike empiricists, Chomskyans have not conducted extensive experiments to confirm (or disconfirm) their theories. This is a serious shortcoming of the Chomskyan science.

In addition the celebrated conceptual move from 'language' as set of sentences to 'language' as part of human brains is highly problematic. Essentially this move has never been fully completed and Chomsky continues to treat language as both: a set of sentences and a biological object (e.g., Chomsky 1986, 1995, 1999, 2000). This treatment blurs the distinction between the object of linguistic study (sentences of a language and their logical relations) and the object of physiological/neurological study (brain structures involved in generating the

sentences linguists can analyze). Chomsky's conflation of the physical tokens of sentences with the non-physical types of sentences results in the untenable view that languages are both finite (as parts of finite brains) and infinite (as grammatical strings of words).

Initially many problems with Chomskyan linguistics were obscured by the rapidly growing complexity of his theories. The frequently changing focus of Chomsky's work led to the introduction (and often subsequent retraction) of countless "rewrite rules, kernel forms and transformations" (Boden, 2006, p. 637), later of principles and parameters, functional chains with multiple applications of move rules, empty categories and much more (Chomsky, 1986). All these, often highly complex, elements were supposed to connect observable surface phenomena and deep structure. All this work resulted in a huge amount of very technical literature but not necessarily in a better understanding of grammar.

Furthermore, the promise to provide mathematically precise descriptions of linguistic phenomena has never been fulfilled. To date we have no computational theory that generates all and only the grammatical sentences of a language. Nor do we have precise descriptions of the formal properties of UG or LAD. According to some critics, this is due not only to the failure of Chomskyan theories but also to the fact that language so defined is not a proper subject of scientific inquiry. In spite of intense research we still lack specific criteria for the boundary between the grammatical and the ungrammatical utterances of any language and "No one has ever successfully produced a comprehensive and accurate grammar of any language" (Graddol, 2004, p.7).

Yet, Chomskyans continue to claim that they have made great progress towards the goal of representing linguistic facts mathematically: "given an I-language specification of a person's lexicon at a time, plus parameter settings, it is possible (in principle) to specify the set of linguistically expressible 'perspectives' that that I-language has to offer" (McGilvray, 2009, p. 47). How can this be?

First, like many Chomskyan claims, this assertion by McGilvray's is too vague to allow evaluation. He does not specify what he means by 'parameter setting' or 'set of linguistically expressible perspectives' or how these elements could be mathematically represented. These complaints about vagueness are not new. Montague (1970) called Chomskyan grammar mathematically imprecise and unsystematic, and Sampson (1979) claimed Chomsky's mathematical argument was often 'maladroit' and sometimes 'just plain wrong'. Others observed "in some [Chomskyan] works we even find purported theorems being stated without any proof being suggested, or theorems that are given 'proofs' that involve no definition of the underlying class of grammars and are thus empty" (Gazdar et al., 1985, p. 14).

Second, many linguists lack the mathematical education that would allow them to evaluate the formal claims made by Chomskyans and to keep up with mathematical implications of the continual recasting of Chomsky's theories. They have to rely on the reasonable assumption that claims made in Chomsky's published work are correct. However, this was often not the case. And in several instances proofs and mathematical formalisms did not even exist (Sampson, 1979). For example, one of Chomsky's core claims about the necessity of transformations was

based "not on a proof, but on a hunch" (Boden, 2006, p. 663).

Given my findings about Chomskyan linguistics, the question may arise if there is anything salvageable. Considering his voluminous output, the potential gain of finding the pieces worth keeping may seem not to justify the tremendous effort required for completing this work. Such a judgment may be premature. But I suggest that the burden of proof is on Chomskyans. They need to show how Chomsky's theories can be stated in unambiguous terms and whether or not the criticisms discussed here can be addressed. Given that Chomsky's work is work in linguistics the evaluation criteria need to be those accepted in linguistics. One suggestion is that grammatical argumentation should meet the following criteria:

> (a) formulate the descriptive devices explicitly,
> (b) consider all previously proposed or 'natural' alternatives to our grammar,
> (c) determine where the alternative grammars differ in empirical consequences,
> (d) compare the alternative grammars with the data and with each other,
> (e) search for crucial examples and arguments to choose one grammar over the others (Dougherty, 1973, p. 457).

It is of course possible to depart from this evaluation matrix. For example one may not need to consider all previously proposed grammars but only those that have not been refuted. But it would seem desirable to provide a thorough evaluation of both: Chomsky's own most recent proposals and the proposals of other linguists.

6.3. Empiricist Views of Language Acquisition

Throughout his career Chomsky has argued for the necessity of an innate

domain-specific LAD. The poverty of the stimulus argument was the main argument that he put forward, allegedly showing that general-purpose learning abilities plus stimuli available to infants are insufficient to account for the accuracy and speed of language learning and the complexity of the language learned (e.g., Chomsky, 1966, 1977, 1988, 2005, 2007, 2009). Chomsky claims that his poverty of the stimulus argument is supported by rich empirical evidence (e.g., Chomsky 1975, 1985, 1994, 2000, 2002).

My survey of recent literature has shown that the empirical evidence accumulated by scientists in various domains (e.g., developmental psychology, experimental and computational language acquisition research) is compatible with a data-driven general-purpose model of language acquisition and does not support the conclusion that an innate domain specific LAD is *necessary* to account for language acquisition. I believe that Chomsky has not fully accounted for several factors that are crucial for language acquisition. For example, he seemingly underestimates the power of general-purpose learning processes and does not account adequately for the incremental nature of the initial process of language learning. Further, recent studies have shown that early language stimuli contain abundant statistical information, perhaps sufficient for language learning given the kind of domain-general learning mechanism children use in other cognitive domains.

I have discussed literature showing that to date extensive domains of language acquisition and input data remain unstudied. We have learned more about the vast quantity of utterances children are exposed to, but we still have neither complete input nor production data for individual children. To adequately evaluate

324

poverty of the stimulus arguments we need much more empirical work in order to understand what sorts of utterances constitute the typical input to children (Pullum & Scholz, 2002; Tomasello, 2003). Thus, at this time the empirical evidence does not support the strong conclusions drawn by Chomsky.

Next I showed that we need to pay attention to the all the steps that are involved in early language acquisition (Tomasello, 2000). During infancy the child learns crucial information about language that will help her in the acquisition of more complex structures later on. Many of the cognitive processes involved in early language acquisition could be based on domain-general learning mechanisms. It is true that children acquire language earlier and seemingly with greater ease than any other abstract domain. However, they learn a considerable amount of information about the nature and organization of their native language long before they begin to produce recognizable words. Language learning appears to be rather slow in the first 18 months of life if we measure it solely by vocabulary acquisition, and it takes a considerable amount of time before children produce grammatically correct sentences. Then the production seems to be for some time confined to short familiar frames, and it takes time again until children produce sentences that are both novel and grammatically complex.

Regarding the innateness debates, it is important to account for all learning that occurs. It is still difficult to evaluate exactly when language learning begins. The research I discussed showed that there is considerable language input and language learning during the prenatal and early infancy periods. Exposure to spoken language begins when the child is still *in utero*. Most fetuses begin to respond to

sound at 22 to 24 weeks, and by the time babies are born their basic auditory capabilities are relatively mature. At birth infants are able to discriminate the voice of their mother from those of other women and seem to be familiar with rhythmic properties of their native language. Several seemingly innate preferences can be explained on the basis of learning from pre-natal input (see Moon & Fifer, 2000).

Further, it also became evident that it takes the infant several months to acquire abilities that are often considered to be innately available. For example, during the first year of life infants learn step by step to fine-tune the perception of the individual sounds that distinguish between words (or phonemes) in the language to which they are exposed. Initially babies distinguish between a wide range of sound differences that signal changes in meaning either in their native language or in non-native languages. But by the time they are 10- to 12-months old infants discriminate only native contrasts. It has been suggested that these changes in perception reflect the growing ability of the infant to focus her attention only on those acoustic dimensions that are relevant for her native language. Findings like these suggest that language learning begins a lot earlier than often suggested, and that a considerable amount of learning takes place during the first year of life. Thus, we need to re-evaluate carefully how many seemingly innate abilities are indeed 'genetically predetermined'.

Furthermore, we also need to account for all the information that can be extracted from the language input. By now we have also learned that statistical regularities of language provide a wealth of implicit information to the young language learner. Young infants already can track the statistical regularities of

language, and they can use this statistical information in early language learning tasks. Researchers have shown that young infants can use acoustic and phonological characteristics to distinguish between function words and content words. Across the world's languages, content words are more salient both acoustically and phonologically. They tend to be longer, to have fuller vowels, and to have more complex syllable structure than function words (Shi et al., 1998). As infants get older, they seem to develop a preferential bias for listening to content words. It has been suggested that this preference is based on the acoustic and phonological salience of content words. The emerging preference for content words could lead to a listening bias that allows infants to learn more about content words, and to use this knowledge in subsequent language acquisition tasks.

Many recent studies suggest that infants integrate different cues in language learning. It might be possible that a combination of several cues reduces the complex task at hand to simpler components. Accessing the multiple sources of information contained in the language input via a combination of several general-purpose mechanisms (acoustic perception, tracking of statistical regularities, etc.) might have the same effect as one highly complex faculty specific mechanism (such as an LAD). More research is needed to establish whether or not language acquisition can be accounted for by data-driven general-purpose learning.

In addition to holding that the language input is impoverished, Chomsky also has suggested that language learning follows certain, genetically predetermined stages (e.g., Chomsky 1966, 1975, 1980, 1986, 2000, 2005). For example, the fast acquisition of vocabulary (vocabulary spurt) and syntax after the second birthday is

frequently used as supporting evidence for the existence of language-specific learning mechanisms that mature at genetically predetermined times (e.g., Chomsky, 1975, 1985; Lightfoot, 1989; Pinker, 1994; Smith, 1999). However, recent research suggests an alternative account. McMurray (2007) conducted a series of simple computational simulations of vocabulary learning. Assuming that (i) words are learned in parallel and (ii) words vary in difficulty, he developed a model that replicated the vocabulary spurt. On his view the vocabulary spurt is an inevitable result of the infant's immersion in words of varying difficulty. Further, Ganger and Brent (2004) found that only one out of five children goes through a verifiable vocabulary spurt. Findings like these suggest that we must rethink whether the vocabulary spurt provides supporting evidence for the existence of the genetically determined maturation of an innate language faculty that is shared by all members of the human species.

According to Chomskyans, positive evidence alone will always be insufficient for language acquisition. Thus, even if it turns out that the positive evidence is considerably richer than previously thought and that general-purpose learning mechanisms can extract all this formation from the input, children still need powerful innate constraints to learn language. This is because, allegedly, negative evidence is needed to allow children to recover from overgeneralization errors, and, allegedly, this negative evidence is missing and/or ignored. It was widely accepted that Gold (1967) had formally proven that it is impossible to learn natural languages without negative evidence. My discussion of the literature has revealed several challenges to the arguments for the necessity of negative evidence.

First, it has been suggested that the 'goal' of language learning assumed by Gold was too ambitious (Chater and Vitanyi, 2007). Gold's proof concerned ideal learners who achieve an end state of 'grammatically perfect' performance. Human learners only achieve an end state that approximates to a higher or lesser degree grammatically perfect performance. The researchers showed that in order to achieve *this* far more modest goal, negative evidence is not necessary. Therefore, Gold's proof is not relevant to actual language acquisition under 'real world conditions' conditions, aiming at 'real world outcomes'. Under these conditions it is at least possible that natural languages can be learned from positive evidence alone.

Second, it has been questioned whether negative evidence is in fact inaccessible to the language learner. Research has shown that different forms of negative evidence are explicitly (e.g., correction of errors) and implicitly (e.g., non-occurrence of expected forms) available to the child. It has been argued that because negative evidence is not given consistently from all parents nor in all instances of ungrammatical utterances of a given child, it is not sufficient for learning from the input alone (e.g., Marcus, 1993). In response to these arguments it has been suggested that from the fact that the available negative evidence is not sufficient for language acquisition it does not follow that this evidence is not useful. When considered in the context of *all* evidence that is available from the input, negative evidence can assist language learning. Thus, the example of negative evidence shows one more time that careful empirical analysis does not confirm strongly held nativist intuitions.

My survey of empirical work suggests that at this point the available

evidence does not support many of the categorical claims put forward by proponents of the poverty of the stimulus argument. Over time, children learn to integrate information from various sources into a coherent whole. Some of the strategies they use in language learning are also employed in non-language related cognitive tasks and by non-human species. In order to maintain the claim that a species-specific LAD is necessary, it needs to be shown that learning strategies that are shared by non-human species cannot account for all aspects of language learning.

I have focused mainly on the learning processes in infants because I want to stress that learning begins very early and that children bring many acquired and not necessarily language-specific abilities to the table when they begin to tackle the syntactically complex structures that are the main focus of Chomsky's work. A comprehensive evaluation of the LAD hypothesis needs to include all of the steps involved in language acquisition. Obviously, the empirical work discussed here cannot rule out the LAD hypothesis. But the results obtained can direct future research. Equally, reformulating the Chomskyan hypotheses in ways suggested in chapter 3 can provide a theoretical framework that can be tested empirically. Much research remains to be done before we know whether children who acquire language use the same general-purpose mechanisms that are exploited in other domains of human cognition, by non-human animals or by computational models.

Given the tremendous amount of data that have been gathered during the last decades, it is crucial to update the theoretical framework frequently. Because the results from current research will determine to some degree future theorizing, it is

important to develop models that implement the current hypothesis in a formal and explicit manner so that it can be tested. In this context philosophers can contribute to conceptual clarity. Given that we still lack satisfactory definitions of key terms like 'innateness', 'language faculty' and 'learning', much conceptual work remains to be done. Ideally this work will be informed by the results of empirical research from all fields of inquiry. Again, these insights are not new. Already Descartes acknowledged the need for careful empirical observation and for interdisciplinary cooperation:

> I know of no other way to discover [scientific truth] than by seeking further observations whose outcomes vary according to which of these ways provides the correct explanation. Moreover, I have now reached the point where I think I can see quite clearly what line we should follow in making most of the observations which serve this purpose: but I also see they are of such kind and so numerous that neither my dexterity nor my income…could suffice for all of them….[therefore] all who desire the general well-being of mankind …[would be obliged] both to communicate to me the observations they have made and to assist me in seeking those which remain to be made. (CSM I, p. 144)

6.4. Computational modeling of language acquisition

In addition to empirical work with children, I evaluated computational models of language-acquisition. In the past four decades computational language acquisition researchers have attempted to challenge the Chomskyan dictum that language acquisition is domain-specific and depends on innate knowledge. Computational work can challenge this dictum in two different ways. First, if it can

be shown that mechanisms that clearly are not involved in human language acquisition simulate human-like performance, then this proves that the mechanism postulated by Chomsky is not necessary for language acquisition. Second, if connectionist and/or other computational models succeed in simulating language acquisition, then this may allow inferences to the nature of the mechanisms that *are* used in human language acquisition.

Chomskyans have challenged the usefulness of computational models, which allegedly "make no contribution to sciences of the mind" (McGilvray, 2009, p. 24). Given that a growing number of researchers use computational models and claim that their findings can provide important insights regarding the mechanisms that underwrite language acquisition, it is important to establish that their work is not flawed in the way suggested by McGilvray.

McGilvray claims that the problems with computational modeling are not merely contingent but necessary because the models used by connectionists (the only computational language researchers he considers) are inadequate: "[Connectionists'] claim that the mind is made up of 'neural nets' is innocuous; it is their claim about the initial state of the net (undifferentiated, approximating Locke's 'blank slate') and their view about how this net gets its 'content' (by training, learning) that place them firmly in the empiricist camp" (McGilvray, 2009, p. 110). He further suggests, "However ingenious the techniques used and no matter how much technology is thrown at prosecuting them, empiricism is wedded to a picture of the mind and of how it gains and uses language and 'content' that does not and …cannot address readily observable facts about human language acquisition and

use" (Ibid., p. 23). To demonstrate the inadequacy of connectionist modeling McGilvray provides only one specific example: "Elman's famous paper - the most quoted in [cognitive science,]... - on learning nested dependencies. Two problems: (1) the method works just as well on crossing dependencies, so doesn't bear on why language near universally has nested but not crossing dependencies. (2) His program works up to depth two, but fails totally on depth three." (Ibid.). According to McGilvray, an additional problem is that "No one finds children subjected to the training procedures for concepts or language explored by connectionists, for example" (Ibid.).

To evaluate these criticisms I needed to establish whether connectionist models start from the Lockean *tabula rasa*, whether they replicate conditions under which children acquire language and whether Elman's models were flawed in the way described. In addition I needed to establish whether connectionist models are the only computational models of language acquisition.

An unrestricted *tabula rasa* view is indeed problematic and has been critiqued by Chomskyans and Empiricists alike. Yet the Chomskyan critiques are often ambiguous, as the following sample of Chomsky's formulations shows. On the one hand we have clear accusations of *tabula rasaism*: "empiricism insists that the brain is a *tabula rasa*, empty, unstructured, uniform at least as far as cognitive structure is concerned" (Chomsky, 1977) and "empiricist currents, that would have us believe that the human mind is empty, a *tabula rasa*" (Chomsky, 1980, p. 270) On the other hand Chomsky denies vehemently that he accused empiricists of *tabula rasaism* by assuring that he has "repeatedly, consistently, and clearly insisted

that all rational approaches to the problem of learning, including 'associationism' and many others...attribute innate structure to the organism" (Chomsky, 1983, p. 310) and that "rationalists and empiricists alike attribute innate structure and principles to the mind" (Chomsky & Katz, 1975, p. 70). A very charitable interpreter of these conflicting passages might hold that for Chomsky the relevant question is not whether connectionist models incorporate innate structure but how much structure is innate. Yet, McGilvray's formulations conflict with this charitable interpretation. Clearly, he accuses connectionists of Lockean *tabula rasaism*. This is the criticism that needed to be addressed by connectionists.

Elman accounted for the fact that models need some inbuilt constraints already in his early work (e.g., Elman, 1990, 1993). And my survey of recent literature found no evidence for *tabula rasaism* in computational modeling. It revealed, instead, that several researchers have explicitly or implicitly rejected completely unconstrained blank slate views of language acquisition (e.g., Hare & Elman, 1995; Elman et al., 1996; Redington & Chater, 1998; MacWhinney, 2000; McDermott, 2001; Solan et al, 2005; Edelman & Waterfall, 2007; Christiansen & Chater, 2008; Chater & Christiansen, 2009). Thus, the first point of McGilvray's critique does not apply to any existing computational models. Further, it also does not apply to the early connectionist nets, because there was never a genuine blank slate model. The structure of any connectionist network's architecture itself provides a very real and powerful constraint on what can be learned. "During the early days, some connectionists failed to appreciate this, and there were claims that implied that networks were basically *tabulae rasae*. But this represented a

misunderstanding on our part" (Elman, p.c.). This means that McGilvray's critique not only is outdated but also never had a factual base.

The next Chomskyan criticism of computational modeling alleges that the models are irrelevant, because they do not resemble the conditions under which children acquire language. Again it is not easy to find out what exactly the nature of the criticism is. For example, one Chomskyan claims that connectionist models cannot possibly simulate what children do because "it takes a quarter of a million repetitions to train one to recognize *Boys like girls* as well formed" (Smith, 1999, p. 132). While the criticism is not aimed at a clearly identified model (Smith provides no citation), it seems to indicate that Smith objects to the amount of repetition that is needed before a model simulates an aspect of language acquisition. Another Chomskyan alleges, "no one finds children subjected to the training procedures for concepts or language explored by connectionists" (McGilvray, 2009, p. 23). Here we have no specific criticism but rather a very general categorical claim. McGilvray offers no specific examples in support of his allegation, which could indicate that he considers connectionist work flawed on *a priori* grounds. Another criticism is similarly broad. McGilvray claims that connectionist modeling is based on arbitrary pairings of "input to outputs that suit the experimenter's criteria for correct behaviour" (McGilvray, 2009, p. 22-23). Here the concern seems to be that the input given to the models is restricted in ways that the input available to children is not.

My evaluation of available connectionist and other computational models did not confirm these broad criticisms. I easily found evidence suggesting that

researchers are paying close attention to the conditions under which children acquire language. Especially, the CHILDES database provides a rich resource for computational modeling and is used frequently. Researchers can rely on samples collected from children's language acquisition. Many computational language researchers use these samples as input for their models. Further, I have shown how models have developed over time and how newer models incorporate insights from previous models. It has become evident that several aspects of language acquisition have been successfully modeled by now. This is an important finding because it casts doubt on the claim that language can be learned only by mechanisms that have evolved in humans and are both species- and domain-specific. Further, it leaves open the possibility that language can be acquired by mechanisms that are much simpler than previously imagined. Thus, McGilvray's (2009) claim that computational work makes no contribution to the sciences of the mind should be rejected. The question is not whether this work contributes to the sciences of the mind but how it contributes and how these contributions can help to formulate questions for future research. In order to do this, individual models need to be evaluated.

At times Chomskyan criticisms are targeted more directly at individual models. For example, when discussing work on learning nested dependencies, Chomsky claims this work is flawed because "(1) the method works just as well on crossing dependencies, so doesn't bear on why language near universally has nested but not crossing dependencies. (2) [the connectionist model] works up to depth two, but fails totally on depth three" (Chomsky, cited in McGilvray, 2009, p. 24). Again,

the literature review revealed that these criticisms do not apply. Several connectionist and computationalist language-acquisition researchers have replicated the same kinds of processing differences for nesting and crossing dependencies that are observed in children (e.g., Christiansen, 1994; Christiansen & Chater, 1999; Christiansen, Conway, & Curtin, 2005; Christiansen & McDonald, 2009). And none of the models I have discussed failed completely at level three. Again, the performance deterioration reported for computationalist models closely resembles that observed in children (e.g., Elman, 1993, Christiansen, 1994; Christiansen & Chater, 1999; Christiansen, Conway, & Curtin, 2005; Christiansen & McDonald, 2009).

I have shown that the Chomskyan criticisms of computational work often miss the target. But this does not imply that there are no problems with modeling. Researchers are aware of this fact and usually strive to outline the limitations of their models. Conclusions are often tentative and exposed to repeated testing. And, in cases where the results differ from the pre-experimental expectations, these results help to re-design experiments.

6.5. Model Selection and Limitations of Simulation

As philosophers we find ourselves in a difficult position when evaluating the work of linguists, computational scientists or developmental psychologists. As 'outsiders' to these fields we are not always aware of the background assumptions

that support specific models, what the goals of these models are, how much and what kind of preliminary work was completed before the reported experiments were conducted, etc. Thus, on the one hand charity requires that we assume that the experts 'know what they are doing'. On the other hand it is our job as philosophers to question the background assumptions and the justification for choosing one experimental set-up over another.

Depending on the background assumptions about the phenomenon under investigation, researchers will consider different models as useful. I will discuss here first some Chomskyan suggestions regarding appropriate models for language and compare their approach to that taken by empiricists. Chomsky has often suggested that linguistics is a natural science and should be pursued with the same methods as other natural sciences. Regarding the modeling of (aspects of) language he writes,

> [In the 17th and 18th century] the project of machine simulation was actively pursued, but as a way to find out something about the world. The great artificer Jacques de Vaucanson did not seek to fool his audience into believing that his mechanical duck was digesting food, but rather to learn something about living things by construction of models, as is standard in the sciences. Contemporary debate contrasts rather unfavorably with the tradition, it would seem. (Chomsky, 2000, p. 114).

Like many of Chomsky's criticisms, this quote conveys opprobrium without being clear what the deficit of contemporary debates is. There are some possible interpretations. First, Chomsky could suggest that some contemporary modelers are trying to 'fool' their audiences by claiming that their results are relevant to human language when in fact they are not or by pretending that internal language

mechanisms are explored. Some of the critical comments in Chomsky (1986, 2007, 2009) would support this interpretation. I will return to this possibility below.

Second, Chomsky could allege that the efforts of non-Chomskyan modelers are misguided. It is accepted that "Vaucanson's automata were philosophical experiments, attempts to discern which aspects of living creatures could be reproduced in machinery, and to what degree, and what such reproductions might reveal about their natural subjects" (Riskin, 2003, p. 601). By contrast, Chomsky might suggest that some of the contemporary models focus on irrelevant aspects of language. This interpretation is supported by his insistence that the only way to learn something meaningful about language is to study I-language (e.g., Chomsky 1980, 1986, 2007). On this view the study of E-language will not reveal anything interesting about language. If this is the criticism, it implies that many contemporary computational models of language acquisition and language processing focus on what he calls an unimportant epiphenomenon (E-language, language behaviour, language performance etc.). To use an analogy: Chomsky seems to think that the computational work is comparable to studying the steam clouds escaping from a steam engine, when we are trying to obtain knowledge about the internal mechanisms of a steam engine. We may learn in some roundabout way something about steam engines that way. But given that the shape of the steam clouds is also affected by weather conditions, traveling speed, etc., our 'findings' will not be very useful. By contrast, Chomsky seems to think, that when Vaucanson built the mechanical digesting duck, he focused on the internal mechanisms that were essential for digestion. Therefore, his model could actually teach us something about duck digestion.

It is of course a matter of debate whether the steam-engine analogy is justified. Assuming for argument's sake that it is, we should reasonably expect that all Chomskyan models of language acquisition resemble the good qualities of the Vaucansonian Duck because Chomsky's research focuses on the essence of the human language faculty. Yet, it is not easy to find such models. Early in his career Chomsky stressed the fact that a generative grammar is not itself to be regarded as a model of the speaker but that successful models would be expected eventually to incorporate generative grammars (e.g., Chomsky, 1965). But to date seemingly no such 'successful model' has been developed. Much of Chomsky's work has focused on sentence analysis and the refinement of a formal descriptive apparatus of ever-increasing complexity (Smith, 1999). The sentences used in Chomskyan analysis are often not taken from actual speech samples but esoteric constructions that "are unlikely to occur in everyday speech" (Smith, 1999, p. 33). Chomskyans have attempted to justify the lack of models that incorporate representative samples of actual speech by reference to the complexity of the phenomenon under investigation and to similar practices in other areas of natural science:

> There is no requirement in science that the phenomena used as evidence should occur spontaneously under natural conditions. The major findings of particle physics have been based on hugely complex experiments carried out using more and more powerful (and expensive) accelerators. To make progress in such a field it is necessary to study extreme conditions. *Similarly*, to understand language it is sometimes necessary to study extreme examples. (Smith, 1999, pp. 33-4)

Smith is correct to assert that *some* findings in *some* areas of physics rely on conditions that do not occur spontaneously. However, he has not shown that and/or to

what degree natural-language research is similar to particle physics. Many foundational findings in biology, the science that allegedly covers language, *are* based on investigation of representative samples that are obtained under 'normal' conditions. So, if language is a biological organ like other biological organs (e.g., Chomsky, 1986, 2000, 2002, 2010), we should expect that language would be investigated with similar methods to those for other biological organs. If language is as dissimilar from other biological organs as particle physics is from Newtonian physics, we should expect that Chomskyans would provide a description of the relevant differences. To my knowledge no such description exists. Therefore, it would seem prudent to remain more Cartesian about the selection of the phenomena we want to study: "Before showing that we can adapt nature to operations that are quite out of the ordinary, we must explain the laws of nature and the ordinary workings of nature" (CSMK III, p. 92). Thus, before tackling esoteric samples, we should begin with ordinary samples when we want to explain the ordinary workings of the language faculty.

Another widely accepted method in modeling aspects of human biology is using animal models. For example, the use of animal models in the health sciences allows the investigation of certain systems of human biology, which are not accessible to direct experimentation in humans. Of course, the results achieved from those models can only serve as hypotheses about how the target system works. Ideally the hypotheses can lead to predictions that can be experimentally verified in those target systems. In a recent publication by a Chomskyan, such an animal model is discussed. Charles Gallistel describes the path-integration organ that allows bees to

orient themselves in space in relation to the sun (for details see Gallistel, 2010, pp. 195-201). This organ consists basically of an internal circadian clock and a compass direction detector. It uses path-integration computation to come up with an ephemeris function for the position of the sun during the day. Gallistel claims that there is a profound poverty of stimulus because the bees need to see the sun only a few times (as opposed to at least one complete 24 hour cycle) to calibrate this mechanism. For this reason Gallistel argues that most of this mechanism must be innately fixed. He believes this is a good model for language acquisition, and this is how he thinks the analogy to language works:

> This mechanism leaves two things to be specified: by the bee's experience [1] the terrain views associated with a given set of firing rates...and [2] the kinetics of the change in the firing rates, that is how they change as the circadian-clock mechanism runs through its cycle. The terrain views are analogous to the words in a lexicon. The kinetics of the ephemeris function are analogous to the grammar of a language. Observing the sun a few different times of day suffices to determine the value of the free parameters in the innately specified ephemeris function. Fixing the values of these parameters adapts the universal function to the locally appropriate variant (Gallistel, 2010, p. 201)

It is possible that the analogies Gallistel suggests hold on some level of analysis. But I think it is questionable that this model gives us a better understanding of the human language faculty. Gallistel seems to suggest that vocabulary learning for us is just like learning the location of a few landmarks is for the bee. But he does not elaborate *how* this learning occurs in the bee's case or why it is similar to vocabulary learning. And we also have only Gallistel's assertion *that* the kinetics of the

ephemeris function and the grammar of a human language are similar. But, apart from the fact that supposedly in both cases a 'poverty of the stimulus' situation occurs, we are not told how the mechanisms in the bee brain resemble those in the human brain. Given that Chomsky has for decades insisted that human language is unlike anything else found in the animal kingdom, we cannot assume that, when learning language, the human brain just does a lot more of what the bee brain does. Thus, Gallistel's analogies do not provide any information we did not already have before he introduced his model.

In fact, the bee analogy bears a quite unfortunate resemblance to an aspect of Vaucanson's Duck I have not mentioned before. While it appeared as if the duck was eating and digesting grain

> …the food did not continue down the neck and into the stomach but rather stayed at the base of the mouth tube. … the grain input and excrement output were entirely unrelated and … the tail end of the Duck must be loaded before each act with fake excrement. The Duck that pioneered physiological simulation was, at its core, fraudulent. Yet, this central fraud was surrounded by plenty of genuine imitation. (Riskin, 2003, p. 609)

Chomsky may be correct to assert that Vaucanson did not merely intend to deceive his audience. Many features of the mechanical duck resembled closely features of real ducks. But if Vaucanson's intention was "to learn something about living things by the construction of models" (Chomsky, 2000, p. 114), then it is peculiar that he would not even attempt to model the process he allegedly studied (digestion). And it would seem that Gallistel's bee model of language is analogous to Vaucanson's Duck because it also leaves out exactly the part of the mechanism

343

that is in need of explanation. As far as we know, it is questionable whether or not the bee model contains at least the same amount of 'genuine imitation' of non-essential mechanisms as the mechanical duck. Thus, I would suggest that the bee model tells us as little about the mechanisms involved in language as the duck model told Vaucanson's contemporaries about the mechanism of digestion. Having found no evidence for the existence of the 'successful models' Chomsky promised 55 years ago, I suggest we ought to take seriously the competing models of language acquisition and language processing that have been dismissed by Chomskyans.

For "those with an *appreciation* of the history of the discipline" (Chomsky, 2000, p. 48, emphasis added) it will be informative to return again to the 17th century and look at some other models crafted by Vaucanson: the flute player and the pipe-and-tabor player. Unlike the duck these artificial musicians did not involve any deception but replicated with considerable accuracy human performance. The flute player was an elaborated mechanism that played a real flute. He was 'blowing' air from mechanical lungs and moving soft, flexible mechanical fingers, lips, and tongue. The design of the flute player was based on careful observation. "Vaucanson studied human flute players in minute detail. He devised various ways of transmitting aspects of their playing into the design of his android" (Riskin, 2003, p. 614). This careful study of the phenomenon under observation is one important foundation of scientific modeling. Another important provision is the formulation of a theory or hypothesis that can be tested by the model. Again, we can see that Vaucanson pioneered this approach when he built the flute player:

> …[Vaucanson devised] a theory of the physics of sound production in the flute, the first known such theory. Vaucanson's idea was that the pitch of a note depended upon the speed of the air's vibrations as it left the flute. This in turn depended upon three parameters: blowing-pressure, the shape of the aperture, and the sounding-length of the flute damping the vibrations, which was determined by the player's finger positions. Vaucanson wanted to test the influence of these three parameters on pitch, and his Flute-player was an acoustical experiment. (Riskin, 2003, p. 615)

Thus, the careful preparation allowed Vaucanson to build a model that closely simulated the performance of a human flute player. But this was not all. The model also confirmed Vaucanson's theory of sound production. This is noteworthy, because at the time the proposal that blowing pressure affected pitch was "in conflict with the recommendations of some contemporary published flute tutors [who] denied that pitch was controlled by blowing-pressure" (Riskin, 2003, p. 616).

As we have seen in chapter 5 some of the computational work on language acquisition has also led to the corrections of pre-modeling beliefs. This work could then be seen as similar in spirit to Vaucanson's flute-player 'modeling'. Computational language acquisition researchers use child-directed speech as input for their models, devise a theoretical framework to be tested and adjust their theories, should the results suggest such adjustments are required. And, as in the earlier modeling attempts, there are limitations to contemporary models. For this reason we need to inquire whether and how researchers in the field acknowledge the limitation of their models. Fortunately, we do not have to look far. Many researchers make explicit comments about what they intend to achieve with a given model. For example Elman writes:

This simulation should not be taken as a model of word acquisition. While listeners are clearly able to make predictions based upon partial input (Grosjean, 1980; Marslen-Wilson & Tyler, 1980; Salasoo & Pisoni, 1985), prediction is not the major goal of the language learner. Furthermore, the co-occurrence of sounds is only part of what identifies a word as such. The environment in which those sounds are uttered, and the linguistic context, are equally critical in establishing the coherence of the sound sequence and associating it with meaning. This simulation focuses only on a limited part of the information available to the language learner. The simulation makes the simple point that there is information in the signal that could serve as a cue to the boundaries of linguistic units, which must be learned, and it demonstrates the ability of simple recurrent networks to extract this information. (Elman, 1990, p. 193).

Thus, far from claiming that his models replicate what goes on in the brains of language-learning children, Elman only asserts that 'information in the signal that could serve as a cue to the boundaries of linguistic units' and that SRNs can extract this information. He claims neither that this information is sufficient for language acquisition nor that children do access this information when *they* learn language. Further, the results obtained from the early models have been used to design successive models (e.g., Elman, 1991, 1993).

Other models are designed to resemble more closely the conditions under which children learn language. Often computational language-acquisition researchers collaborate with experts from other fields. This is important because "collaborative efforts between modelers and experimentalists can obtain a deeper understanding of how infants [acquire different aspects of language]" (Blanchard et al., 2010, p. 488). Incidentally, this interdisciplinary approach can be traced to some Cartesian roots: "all the sciences are so closely interconnected that it is much easier

to learn them all together than to separate them. If …someone…wishes to investigate the truth of things he ought not to select one science in particular, for they are all interconnected and interdependent" (CSM I, p. 10). In these interdisciplinary collaborations it is very important that the researchers clearly outline the limitations of their models and the expectations that motivated them to make specific modeling choices. Below the interaction between findings in developmental psychology and computational modeling is discussed:

> …the process of segmentation itself undergoes change with development as more cues are discovered by the infant. Each of these tenets makes empirical predictions about the developmental course of segmentation. The virtue of modeling segmentation is that it helps us understand why certain cues fall out or emerge from the use of earlier cues, potentially explaining earlier experimental results, as well as suggesting further experiments to test predictions the model makes. (Blanchard et al., 2010, p. 494).

Here the focus is clearly on developing models that closely mirror what happens when children learn language. In order to do that successfully, researchers need a thorough understanding of the work completed in developmental psychology. Thus, after work in developmental psychology had shown that children integrate different cues when they begin to segment the words of a language it was possible to design models that simulate this aspect of language acquisition:

> …it is also important to look at the fundamental properties of the models in relation to the task at hand. In the case of modeling infant language segmentation, it is important for the model to utilize the types of cues that infants do, and combine them in a way that does not conflict with data on how infants process language. (Blanchard et al., 2010, p.

507)

It would seem that the focus on the conditions under which children acquire language is a promising strategy for successful modeling. Interestingly there is also another parallel to Vaucanson's work. He was not only a gifted craftsman but also a versatile inventor. He is credited with the first-ever use of punch cards for the operation of weaving machinery. For this achievement, he is sometimes considered to be the father of modern computer programming. In a way then, the work of computational language researchers can be seen as continuation of this aspect of Vaucanson's pioneering work. Seen from this perspective, not all contributions from the Cartesian period have been ignored by non-Chomskyan researchers.

As the previous discussion has demonstrated, we cannot answer the question 'which model is the best model?' based on *a priori* assumptions or abstract theorizing. When we design and evaluate models, we need to pay close attention to contingent, empirical facts about the phenomenon we want to study. This approach can be traced back at least to Descartes who was acutely aware of the shortcoming of relying only on abstract reasoning which was not informed by empirical evidence.

> Many people...examine difficult problems in a very disorderly manner, behaving in my view as if they were trying to get from the bottom to the top of a building at one bound, spurning or failing to notice the stairs designed for that purpose...This applies also to those philosophers who take no account of experience and think that truth will spring from their brains like Minerva from the head of Jupiter. (CSM I, pp. 20-21)

Descartes makes two important points here. First, if we want to gain a thorough understanding of the phenomenon under investigation we need to proceed in a slow, orderly fashion. Initial models need to be reevaluated carefully because they provide the stepping-stones that lead eventually to the 'top of the building'. Second, we need to 'take experience into account'. That is we need to be open to the possibility that empirical findings can challenge our theoretical framework. For these reasons we should not reject any models based on a priori assumption. The performance of individual models needs to be evaluated carefully with a keen eye on what the model was supposed to simulate and whether or not this task has been accomplished.

On the other hand some general philosophical points about what we should *not* expect from models can be made without evaluating specific models. First, we cannot expect that a model resembles every aspect of the phenomenon that it is a model of. This much we have known at least since Plato:

> … the image, if expressing in every point the entire reality, would no longer be an image. Let us suppose the existence of two objects: one of them shall be Cratylus, and the other the image of Cratylus; and we will suppose, further, that some God makes not only a representation such as a painter would make of your outward form and colour, but also creates an inward organization like yours, having the same warmth and softness; and into this infuses motion, and soul, and mind, such as you have, and in a word copies all your qualities, and places them by you in another form; would you say that this was Cratylus and the image of Cratylus, or that there were two Cratyluses. (Cratylus, 432b/c)

We can easily substitute 'model' for 'image' and will appreciate that it would

be pointless to create a model of language acquisition that resembles every aspect of language acquisition in children. Such a model would no longer be a model but another child. In the *Optics* Descartes developed a model of human vision that was based on an analogy to the imagery used in engravings. Like Plato, Descartes used this analogy to point out that it is not informative to create images that resemble every aspect of reality:

> … in no case does an image have to resemble the object it represents in all respects, for otherwise there would be no distinction between the object and its image. It is enough that the image resembles the object in a few respects…in the case of engravings: consisting simply of a little ink placed here and there on a piece of paper they represent to us forests, town, people, and even battles and storms…Moreover, in accordance with the rules of perspective they often represent circles by ovals better than by other circles, squares by rhombuses better than by other squares, and similarly for other shapes. Thus it often happens that in order to be more perfect as an image and to represent an object better, an engraving ought not to resemble it. Now we must think of the images formed in our brain just in the same way, and note that the problem is to know simply how they can enable the soul to have sensory perceptions of all the various qualities of the objects to which they correspond. (CSM I, pp. 165-166)

Descartes makes several points here. Like Plato, he points out that any model will by necessity leave out certain aspects of reality. Furthermore, some aspects of models will not resemble accurately the corresponding properties of the objects that are modeled. Finally, it is essential that the models allow us to focus on the aspects of realty that are in need of explanation. Thus, a good model of human vision will be one that allows us to gain a better understanding of how images on the retina 'enable the soul to have sensory perceptions of all the various qualities of the objects'. And, presumably, a good model of language acquisition is one that allows us to gain a

better understanding of how the interaction of language input and internal resources enables the child to learn language.

While scientific models have been refined in many ways since Descartes' time, his fundamental insights still apply. Models are not complete replicas of reality. They can only focus our attention on certain aspects of reality. Thus, it is always important to be clear about which aspects are left out and why. Knowing this will allow us to evaluate the models better. As we have seen in chapter 5, we cannot rely on *a priori* assumptions to judge models, but we need to evaluate specific models carefully.

Language-acquisition models will in some degree always depend on what is already known about language acquisition. And, at least currently, models can only simulate small, well-defined aspects of the language-acquisition task. These aspects do not occur in isolation when children learn language. Therefore, we need to remain cautious about the interpretation of the results from modeling. Last but not least, we need to remember that 'performance similarities' between models and children do not necessarily indicate that the underlying mechanisms are the same in both cases. Again, Descartes already cautioned against our propensity to perceive similarities: "Whenever people notice some similarity between two things, they are in the habit of ascribing to the one what they find in the other, even when the two are not in that respect similar" (CSM I, p. 9). Whether or not our aim is to construct a *Cartesian Linguistics*, we are well advised to pay attention to Descartes' insights.

Bibliography

Aarsleff, H. (1970). The History of Linguistics and Professor Chomsky. *Language, 46,* 570-585.

Aarsleff, H. (1971). 'Cartesian Linguistics': History or Phantasy? *Language Sciences, 17,* 1-12.

Abel, R. (1968). On the alleged uniqueness of language. In: Hook, S. (Ed.) *Language and Philosophy.* (pp. 120 - 121). New York: New York University Press.

Abelson, R. (1968). Depth Grammar and Necessary Truth. In: Hook, S. (Ed.) *Language and Philosophy.* pp. 215-220. New York: New York University Press.

Adams, R. (1975). Where do our ideas come from? - Descartes vs. Locke. In: Stich, S. (Ed.) *Innate ideas.* pp. 71-87. Berkeley: University of California Press.

Akhtar, N., Jipson, J., & Callanan, M. (2001). Learning words through overhearing. *Child Development*, *72*, 416-430.

Alanen, L. (2003). *Descartes' concept of mind.* Cambridge: Harvard University Press.

Alanen, L. (1994). Sensory ideas, objective reality, and material falsity. In: Cottingham, J. (Ed.) *Reason, Will, and Sensation: Studies in Descartes's Metaphysics.* pp. 229 - 250. London: Clarendon Press.

Allen, J. & Van Buren, P. (Eds.). (1971). *Chomsky: Selected Readings* London: Oxford University Press.

Allen, J. & Seidenberg, M. (1999). The emergence of grammaticality in connectionist networks. In: B. MacWhinney (Ed.). *The Emeregence of Language.* (pp. 115-152). Mahwah, NJ: Lawrence Erlbaum Associates.

Anderson, J. (2010). Historiography's contribution to theoretical linguistics. In: Kibbee, D. (Ed.). *Chomskyan (R)evolutions.* (pp. 445 − 471). Amsterdam: John Benjamins.

Anderson, S. & Lightfoot, D. (2002). *The language organ: Linguistics as cognitive physiology.* Cambridge: Cambridge University Press.

Ariew, R., Des Chene, D., Jesseph, D., Schmaltz, T., & Verbeek, T. (2003). *Historical Dictionary of Descartes and Cartesian Philosophy.* Oxford: The Scarecrow Press.

Ashworth, E. (1972). Descartes' theory of clear and distinct ideas. In: Buttler, R. (Ed.) *Cartesian Studies*. Oxford: Basil Blackwell.

Ashworth, E. (1975). Descartes' theory of objective reality. *The New Scholasticism, 49,* 331-340.

Baker, G., & Morris, K. (1996). *Descartes' Dualism.* London: Routledge.

Baker, C. (1979). Syntactic theory and the projection problem. *Linguistic Inquiry, 10*, 533–581.

Baker, M. (2003). *The Atoms of Language: The Mind's Hidden Rules of Grammar.* Oxford: Oxford University Press.

Baker, M. (2005). The Innate Endowment for Language. In: S. Laurence, P. Carruthers, & S. Stich (Eds.), *The Innate Mind: Structure and Contents*. (pp. 156-174). Oxford: Oxford University Press.

Bandura, A. (1977). *Social learning theory.* Englewood Cliffs, NJ: Prentice Hall.

Barnes, J. (1972). Mr. Locke's Darling Notion. *Philosophical Quarterly, 22*, 193-214.

Barsky, R. (1997). *Noam Chomsky: A life of discent.* Cambridge, MA: MIT Press.

Bateson, P. (1991). Are there principles of behavioural development? In: Bateson (Ed.), *The Development and Integration of Behaviour: Essays in honour of Robert Hinde*. (pp. 19–39). Cambridge: Cambridge University Press.

Beedham, C. (2010). The equivocation of form and notation in generative grammar. In: Kibbee, D. (Ed.) *Chomskyan (R)evolutions.* (pp. 19 – 42). Amsterdam: John Benjamins.

Behan, D. (2000). Descartes and formal signs. In: Gaukroger, S., Schuster, J., & Sutton, J. (Eds.) *Descartes' natural philosophy.* pp. 528 - 541. London: Routledge.

Behme, C. (2011). *Cartesian Innate Ideas and Chomskyan Linguistics.* Presented at 55th Annual Congress Canadian Philosophical Association. University of New Brunswick, Fredricton, June 1.

Behme, C. (2010). How Cartesian is *Cartesian Linguistics*? In: Hannah, P. *An Anthology of Philosophical Studies*. (pp. 79-88). Athens (Greece): ATINER.

Behme, C. (2009). Review of *Cartesian Linguistcs* by Noam Chomsky. Cambridge University Press. In: *Metapsychology, Volume 13, Issue 36*

Behme, C. & Deacon, H. (2008). Language Learning in Infancy: Does the Empirical Evidence Support a Domain Specific Language Acquisition Device? *Philosophical Psychology 21,* 641 – 671.

Behne, T., Carpenter, M., & Tomasello, M. (2005). One-year-olds comprehend the communicative intentions behind gestures in a hiding game. *Developmental Science, 8*, 492 - 499.

Behrens, H. (2006). The Input-Output Relationship in First Language Acquisition. *Language and Cognitive Processes, 21*, 2-24.

Blanchard, D., Heinz, J., & Golinkoff, R. (2010). Modeling the contribution of phonotactic cues to the problem of word segmentation. *Journal of Child Language, 37,* 487 - 511.

Bloom, P. (2000). *How children learn the meanings of words.* Cambridge, MA. MIT Press.

Bloom P. (2004). Children think before they speak. *Nature, 430,* 410-411.

Bloom, P. (2005). Word learning, intentions, and discourse. *Journal of the Learning Sciences, 14,* 311-314.

Bloomfield, L. (1926). A set of postulates for the science of language. *Language 2,* 153-164.

Bloomfield, L. (1933). *Language.* New York: Henry Holt.

Bod., R. (2009). From Exemplar to Grammar: A Probabilistic Analogy-based Model of Language Learning. Cognitive Science, 33, In press.

Boden, M. (2006). *Mind as machine.* Oxford: Oxford University Press.

Boeckx., C. (2009). Cartesian Biolinguistics. Presented at The State of the Art in Cartesian Linguistics; *A lecture series of the Sophia Linguistic Institute for International Communication (SOLIFIC).* Tokyo.

Boeckx, C. (In press). Some reflections on Darwin's Problem in the context of Cartesian Biolinguistics. In Di Sciullo, M. & Boeckx, C. (Eds.). *The biolinguistic enterprise: New perspectives on the evolution and nature of the human language faculty.* Oxford: Oxford University Press.

Boeckx, C. & Hornstein, N. (2010). The Varying Aims of Linguistic Theory. In: Bricmont, J. & Franck, J. (Eds.) *Chomsky Notebook.* (pp. 109-133). New York: Columbia University Press.

Bohannon, J., & Stanowicz, L. (1988). The issue of negative evidence: Adult responses to children's language errors. *Developmental Psychology, 24,* 684–689.

Bohannon, J., Padgett, R., & Nelson, K. (1996) Useful evidence on negative evidence. *Developmental Psychology, 32*, 551-555.

Bolton, M. (1986). Confused and obscure ideas of sense. In: Oksenberg Rorty, A (Ed.) *Essays on Descartes' Meditations.* pp. 389-403. Berkeley: University of California Press.

Botha, R. (1989). *Challenging Chomsky*. Oxford: Blackwell.

Botvinick, M. & Plaut, D. (2004). Doing without schema hierarchies: A recurrent connectionist approach to normal and impaired routine sequential action. Psychological Review, 111, 395-429.

Boysson-Bardies B., & Vihman M. (1991). Adaptation to language: Evidence from babbling and first words in four languages. *Language, 67*, 297-319.

Boyle, D. (2000). Descartes on innate ideas. *Modern Schoolman; A Quarterly Journal of Philosophy. 78.* 35-51.

Bracken, H. (1982). Chomsky's Variations on a Theme by Descartes. *History of Philosophy,* 181-192.

Bracken, H. (1984). *Mind and Language: Essays on Descartes and Chomsky*. Dordrecht: Foris Publications.

Bricmont, J. & Franck, J. (2010). Chomsky, France, reason, power. In: Bricmont, J. & Franck, J. (Eds.) *Chomsky Notebook.* (pp. 53-73) New York: Columbia University Press.

Brent, M., & Cartwright, T. (1996). Distributional Regularity and Phonotactic Constraints are Useful for Segmentation. *Cognition, 61,* 93-125.

Brent, M. (1999a). Speech Segmentation and Word Discovery: A Computational Perspective. *Trends in Cognitive Science, 3,* 294-300.

Brent, M. (1999b). An efficient, probabilistically sound algorithm for segmentation and word discovery. *Machine Learning, 34,* 71–105.

Brent, M. & Siskind, J. (2001). The role of exposure to isolated words in early vocabulary development. *Cognition, 81,* 31–44.

Brodsky, P., Waterfall, H., & Edelman, S. (2007). Characterizing Motherese: on the computational structure of child-directed language. Proceedings of the 29th Cognitive Science Society Conference. Austin, TX.

Brook, A., & Stainton, R. (2000). *Knowledge and Mind.* Cambridge, MA: MIT Press.

Brown, D. (2006). *Descartes and the Passionate Mind.* Cambridge: Cambridge University Press.

Bruner, J. (1966). *Toward a Theory of Instruction.* Cambridge, MA: Harvard University Press.

Bunge, M. (2003). *Philosophical Dictionary*. Amherst, NY: Prometheus Books.

Bunge, M. & Ardila, R. (1987). *Philosophy of Psychology.* New York: Springer Verlag.

Burling, R. (2005). *The Talking Ape*. Oxford: Oxford University Press.

Byrne, R. & Russon, A. (1998). Leraning by imitation: A hierarchical approach. *Behavioral and Brain Sciences, 21*, 667-721.

Cameron-Faulkner, T., Lieven, E., & Tomasello, M. (2003). A construction based analysis of child-directed speech. *Cognitive Science, 27,* 843–873.

Cartwright, T. & Brent, M. (1997). Syntactic categorization in early language acquisition: Formalizing the role of distributional analysis. *Cognition, 63,* 121 - 170.

Cattell, R. (2006). *An introduction to mind, consciousness and language.* London: Continuum.

Chappell, V. (1986). The theory of ideas. In: Oksenberg Rorty, A (Ed.) *Essays on Descartes' Meditations*. pp. 177-198. Berkeley: University of California Press.

Chater, N. & Vitányi, P. (under review). A simplicity principle for language acquisition: Re-evaluating what can be learned from positive evidence.

Chater, N. & Vitányi, P. (2007). Ideal learning of natural language: Positive results about learning from positive evidence. *Journal of Mathematical Psychology*, *51*, 135 -163.

Chater, N. & Christiansen, M. (2008). Computational models in psycholinguistics In: R. Sun (Ed.). *Cambridge Handbook of Computational Cognitive Modeling* (pp. 477-504). New York: Cambridge University Press.

Chomsky, N. (1953). Systems of Syntactic Analysis. *Journal of Symbolic Logic, 18,* 242 – 256.

Chomsky, N. (1956). Three Models for the Description of Language. *IRE Transactions on Information Theory*.

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton & Co.

Chomsky, N. (1959). Review of Skinner's Verbal Behavior. *Language, 12*, 89-95.

Chomsky, N. (1964). *Current issues in linguistic theory*. The Hague: Mouton.

Chomsky, N. (1965a). *Syntactic Structures*. London: Mouton.

Chomsky, N. (1965b). *Aspects of the Theory of Language*. Cambridge, MA: MIT Press.

Chomsky, N. (1966). *Cartesian Linguistics. A chapter in the history of Rationalist Thought*. New York: Harper & Row.

Chomsky, N. (1967). Recent contributions to the theory of innate ideas. *Synthese 1,* 1-11.

Chomsky, N. (1968). Linguistics and Philosophy. In: Hook, S. (Ed.) *Language and Philosophy*. (pp. 51 - 94). New York: New York University Press.

Chomsky, N. (1970). Deep structure, surface structue and semantic interpretation. In: Jackobsen, R. & Kawamoto, S. (Eds.). *Studies in gernaral and oriental linguistics*. (pp. 52-91). Tokyo: TEC Company.

Chomsky, N. (1971). Problems of knowledge. In: Allen, J. & Van Buren, P. (Eds.). *Chomsky: Selected Readings London:* Oxford University Press.

Chomsky, N. (1972). *Studies on Semantics in Generative Grammar*. The Hague: Mouton.

Chomsky, 1973). Conditions on transformations. In: Anderson, S. & Kieparsky, P. (Eds.) *A Festschrift for Morris Halle*. (pp. 232-286). New York: Holt, Rinehart and Winston.

Chomsky, N. (1975a). *Reflections on Language*. New York: Pantheon Books.

Chomsky, N. (1975b). *Cartesian Linguistics.* In: Stich, S. (Ed.) *Innate ideas*. pp. 89 - 103. Berkeley: University of California Press.

Chomsky, N. (1975c). Recent contributions to the theory of innate ideas. In: Stich, S. (Ed.) *Innate ideas*. pp. 121 - 131. Berkeley: University of California Press.

Chomsky, N. (1975d). *Questions of form and interpretation.* Lisse: Peter de Ridder Press.

Chomsky, N. (1977). *Essays on Form and Interpretation*. Amsterdam: North Holland.

Chomsky, N. (1980). The Linguistic Approach. In: M. Piatelli-Palmerini. (Ed.), *Language and Learning* (pp.107-130). Cambridge, MA: Harvard University Press.

Chomsky, N. (1981). *Lectures on Government and Binding: The Pisa Lectures*. Berlin: Mouton de Gruyter.

Chomsky, N. (1982). *Noam Chomsky on the Generative Enterprise: A Discussion with Riny Huybregts and Henk van Riemsdijk*. Dordrecht, Holland: Foris Publications.

Chomsky, N. (1983). On cognitive structures and their development. A reply to Piaget. In: Piattelli Palmerini (Ed.) *Language Learning: A debate between Noam Chomsky and Jean Piaget*. (pp. 31-59). Cambridge MA: Harvard University Press.

Chomsky, N. (1985). Knowledge of Language: Its Nature, Origin and Use. In R. Stainton. (Ed.), *Perspectives in the Philosophy of Language* (pp. 3-44). Peterborough. ON: Broadview Press.

Chomsky, N. (1986a). *Knowledge of Language*. New York: Praeger Publishing.

Chomsky, N. (1986b). *Lectures on government and binding*. Dordrecht: Foris Publications.

Chomsky, N. (1988). *Language and Problems of Knowledge*. Cambridge: MIT Press.

Chomsky, N. (1990). On formalization and formal linguistics. *Natural Language and Linguistic theory, 8,* 143 –147.

Chomsky, N. (1991). Linguistics and adjacent fields: a personal view. In: Kasher, A. (Ed.). *The Chomkyan Turn*. (pp. 3 - 25). Cambridge, MA: Blackwell.

Chomsky, N. (1993). *Language and Thought*. Wakefield, RI: Moyer Bell.

Chomsky, N. (1995). *The minimalist program*. Cambridge: MIT Press.Chomsky, N. (1996). *Powers & Prospects.* Boston: South End Press.

Chomsky, N. (1997). History and Theory Construction in Modern Linguistics. In: Chomsky no Brasil / Chomsky in Brazil: Revista de documentação de estudos em lingüistica teórica e aplicada (D. E. L. T. A.), vol. 13 (Número especial).

Chomsky, N. (2000 a). *The Architecture of Language*. Oxford: Oxford University Press.

Chomsky, N. (2000 b). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2002). *On Nature and Language.* Cambridge: Cambridge University Press.

Chomsky, N. (2003). Reply to Gopnik. In: Antony, L. & Hornstein, N. (Eds.) *Chomsky and his critics*. Malden, MA: Blackwell Publishing.

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

Chomsky, N. (2006). *Language and Mind*. Cambridge: Cambridge University Press.

Chomsky, N. (2007a). Symposium on Margaret Boden, Mind as Machine: A History of Cognitive Science. *Artificial Intelligence 171,* 1094-1103.

Chomsky, N. (2007b). Language and thought: Descartes and some reflections on venerable themes. In: Brook, A. (Ed.). *The prehistory of cognitive science*. (pp. 38-66). New York: Palgrave Macmillan.

Chomsky, N. (2009a). Cognition and Language. In: Oezsoy, A. & Nakipoglu, M. (Eds.) Noam Chomsky on Language and Cognition. (pp. 5 - 21). Munich: Lincomb GmBH.

Chomsky, N. (2009b). Comments and answers. In: Oezsoy, A. & Nakipoglu, M. (Eds.) *Noam Chomsky on Language and Cognition*. (pp. 22 - 44). Munich: Lincomb GnBH.

Chomsky, N. (2010a). The mysteries of nature how deeply hidden? In: Bricmont, J. & Franck, J. (Eds.) *Chomsky Notebook.* (pp. 3-33) New York: Columbia University Press.

Chomsky, N. (2010b). An interview with Noam Chomsky. In: Bricmont, J. & Franck, J. (Eds.) *Chomsky Notebook*. (pp. 74-110) New York: Columbia University Press.

Chomsky, N. & Katz, J. (1975) 'On innateness: A reply to Cooper' *The Philosophical Review 84*, 70-86.

Chouinard, M. & Clark, E. (2003). Adult reformulations of child errors as negative evidence. *Journal of Child Language, 30,* 637-669.

Christiansen, M., Dale, R., & Reali, F. (2010). Connectionist explorations of multiple-cue integration in syntax acquisition. In: S. Johnson (Ed.), *Neoconstructivism: The new science of cognitive development.* (pp. 87 -108). New York: Oxford University Press.

Christiansen, M. & MacDonald, M. (2009). A usage based approach to recursion in sentence processing. *Language Learning, 59 (Suppl. 1),* 126-161.

Christiansen, M., & Kirby, S. (Eds.), (2003). *Language Evolution*. Oxford: Oxford University Press.

Christiansen, M., Dale, R., Ellefson, M., & Conway, C. (2002). The role of sequential learning in language evolution: Computational and Experimental Studies. In: A. Cangelosi & D. Parisi (Eds.). *Simulating the evolution of language*. London: Springer.

Christiansen, M., & Ellefson, M. (2002). Linguistic Adaptation without Linguistic Constraints: The Role of Sequential Learning in Language Evolution. In A. Wray, (Ed.), *The Transition to Language* (pp. 335-358). Oxford: Oxford University Press.

Christiansen, M. & Chater, N. (2001). Connectionist psycholinguistics in perspective. In M. Christiansen & N. Chater (Eds.), Connectionist psycholinguistics (pp. 19-75). Westport, CT: Ablex.

Christiansen, M. & Chater, N. (1999). Toward a Connectionist Model of Recursion in Human Linguistic Performance. *Cognitive Science, 23,* 157-205.

Christiansen, M., Allen, J. & Seidenberg, M. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes 13*, 221–68.

Christiansen, M. & Allen, J. (1997). Coping with variation in speech segmentation. In A. Sorace, C. Heycock & R. Shillcock (Eds). Proceedings of the GALA ' 97 conference on language acquisition : Knowledge representation and processing, pp. 327–32. Edinburgh : Edinburgh University Press.

Christiansen, M. & Devlin, J. (1997). Recursive inconsistencies are hard to learn: A connectionist perspective on universal word order correlations. In: *Proceedings of the 19th Annual Cognitive Science Society Conference* (pp. 113-118). Mahwah, NJ: Lawrence Erlbaum Associates.

Clarke, D. (2003). *Descartes' Theory of Mind*. Oxford: Oxford University Press.

Clarke, D. (2006). *Descartes: A Biography*. Cambridge: Cambridge University Press.

Cleermans, A. (1993). *Mechanisms of implicit learning: Connectionist models of sequence processing.* Cambridge: MIT Press.

Cleeremans, A., Servan-Schreiber, D., & McClelland, J. (1989). Finite state automata and simple recurrent networks. *Neural Computation, 1*, 372–381.

Cohen, J. (1970). Review of Sydney Hook (ed.) Language and Philosophy: A Symposium. *Journal of Linguistics, 6*, 134-136.

Costa, M. (1983). What Cartesian ideas are not. *Journal of the History of Philosophy, 21,* 537 - 549.

Cottingham, J. (1976). *Descartes' conversation with Burman.* Oxford: Clarendon Press.

Cottingham, J. (1978). Descartes on Thought. *Philosophical Quarterly, 28*, 208-14.

Cottingham, J. (1986). *Descartes.* Oxford: Blackwell.

Cottingham, J. (1992). Cartesian dualism: theology, metaphysics, and science. In: Cottingham, J. (Ed.) *The Cambridge Companion to Descartes.* pp. 236-257. Cambridge: Cambridge University Press.

Cottingham, J. (1997). "The only sure sign . . ." Descartes on Thought and Language. In: Preston, J. (Ed.), *Thought and Language.* pp. 29-50. Cambridge: Cambridge University Press.

Cottingham, J. (2008). *How to Read Descartes.* London: Granta.

Cowie, F. (1999). *What's within? Nativism reconsidered.* Oxford: Oxford University Press.

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

Crain, S. & P. Pietroski (2001). Nature, nurture and Universal Grammar. *Linguistics and Philosophy* 24, 139-186.

Crain, S., & Pietroski, P. (2002). Why Language Acquisition is a Snap. *Linguistic Review, 19,* 163-83.

Crain, S., Gualmini, A., & Pietroski, P. (2005). Brass Tacks in Linguistic Theory. In S. Laurence, P. Carruthers, & S. Stich (Eds.), *The Innate Mind: structure and contents* (pp.175-197). Oxford: Oxford University Press.

Crain, S., Goro, T. & Minai, U. (2007). Hidden units in child language. In: A. Schalley, A. & Khlentzos, D. (Eds.) *Mental States: Nature, Function and Evolution.* pp. 275-294. John Benjamins.

Culicover, P. (1999). Minimalist architectures. (Review article on R. Jackendoff (1997) The Architecture of the language faculty. Cambridge, MA: MIT Press.) *Journal of Linguistics*, *35*, 137-150.

Culicover, P. & Jackendoff, R. (2006). The simpler syntax hypothesis. *Trends in Cognitive Sciences 10*, 413-18.

Cummins, R. & Harnish, R. (1980). The language faculty and interpretation of linguistics. *Behavioural and Brain Sciences, 3*, 18-19.

Cutler, A. (1993). Phonological cues to open- and closed-class words in the processing of spoken sentences. *Journal of Psycholinguistic Research, 22,* 109-131.
D'Agostino, F. (1986). *Chomsky's System of Ideas*. Oxford: Clarendon Press.

Dabrowka, E. (2004). Language, Mind and Brain. Edinburgh: Edinburgh University Press.

Dale, P. & Goodman, J. (2005) Commonality and Individual Differences in Vocabulary Growth. In: M. Tomasello & D. Slobin (Eds.) *Beyond Nature-Nurture: Essays in Honour of Elizabeth Bates.* (pp. 41-78). Mahwah, NJ: Lawrence Erlbaum Associates.

Dale, R. & Spivey, M. (2005). Categorical Recurrence Analysis of Child Language. *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* (pp. 530-535). Mahwah, NJ: Lawrence Erlbaum Associates.

Dale, R. & Spivey, M. (2006). Unraveling the dyad: Using recurrence analysis to explore patterns of syntactic coordination between children and caregivers in conversation. *Language Learning, 56*, 391-430.

Danto, A. (1968). Semantic vehicles, understanding and innate ideas. In: Hook, S. (Ed.) Language and Philosophy. (pp. 122 -137). New York: New York University Press.

Danto, A. (1978). The representational character of ideas and the problem of the external world. In: Hooker, M. (Ed.) *Descartes: Critical and Interpretative Essays.* pp. 287 - 297. Baltimore: Johns Hopkins University Press.

Dawkins, R. (1986). *The Blind Watchmaker.* New York: Norton.

DeCasper, A., & Fifer, W. (1980). Of human bonding: newborns prefer their mothers' voices. *Science, 208,* 1174-1176.

DeCasper, A., & Spence, M. (1986). Prenatal maternal speech influences on newborns' perception of speech sounds. *Infant Behaviour and Development, 9,* 133-150.

Deacon, T. (1990). Rethinking Mammalian Brain Evolution. *American Zoologist, 30,* 629 - 705.

Deacon, T. (1997). *The Symbolic Species.* New York: W.W. Norton & Company.

Demetras, M., Post, K., & Snow, C. (1986). Feedback to first language learners: the role of repetitions and clarification questions. *Journal of Child Language, 13,* 275-292.

Demopoulos, W. (1990). On Learnability Theory to the Rationalism-Empiricism Controversy. In R. Mathews & W. Demopoulos (Eds.) *Learnability and Linguistic Theory, Studies in Theoretical Psycholinguistics.* (pp. 77-88). Amsterdam: Kluwer Academic Publishers.

Des Chene, D. (2001). *Spirits and Clocks: Machine and Organism in Descartes.* Ithaca: Cornell University Press.

Dickinson, S. (1987). Recursion in development: Support for a biological model of language. *Language and Speech, 30*, 239–249.

Diesendruck, G., Hall, D., & Graham, S. (2006). Children's use of syntactic and pragmatic knowledge in the interpretation of novel adjectives. *Child Development, 77*, 16-30.

Dittmar, M., Abbot-Smith, K., Lieven, E. & Tomasello, M. (2008) German Children's Comprehension of Word Order and Case Marking in Causative Sentences. Child Development, 79, 1152-1167.

Doney, W. (1967). Introduction. In: Doney, W. (Ed.) *Descartes - A collection of critical essays*. pp. 7 - 27. Garden City, NY: Anchor Books.

Dougherty, R. (I973). A survey of linguistic methods and arguments. *Foundations of Language 10.* 423-490.

Doull, F. (2008). Hegel on Descartes. In: Robertson, N., McOuat, G., & Vinci, T. (Eds.). *Descartes and the Modern*. pp. 171-194. Newcastle upon Tyne: Cambridge Scholars Publishing.

Edelman, S. & Waterfall, H. (2007). Behavioral and computational aspects of language and its acquisition, *Physics of Life Reviews 4*. 253-277.

Ellefson, M., & Christiansen, M. (2000). Subjacency Constraints Without Universal Grammar: Evidence from Artificial Language Learning and Connectionist Modeling: In L. R. Gleitman & A. K. Joshi (Eds.), *The Proceedings of the 22nd Annual Conference of the Cognitive Science Society* (pp. 645-50). Mahwah, NJ: Lawrence Erlbaum Associates.

Elman, J. (1990). Finding structure in time. *Cognitive Science, 14*, 179-211.

Elman, J. (1991). Distributed representations, simple recurrent networks, and grammatical structure. Machine Learning, 7, 195-224.

Elman, J. (1993). Learning and development in neural networks: The importance of starting small. *Cognition, 48,* 71-99.

Elman, J., Bates, E., Johnson, M., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. Cambridge, MA: MIT Press.

Elman, J. (1999). The emergence of language: a conspiracy theory. In. MacWhinney, B. (Ed.) *The emergence of language.* pp. 1 - 27. London: Lawrence Erlbaum.

Engelmann, F. & Vasishth, S. (2009). Processing grammatical and ungrammatical center embeddings in English and German: A computational model. Submitted manuscript.

Erion, G. (2001). The Cartesian Test for Automatism. *Minds and Machines 11,* 29–39.

Evans, N. & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Behavioral and Brain Sciences, 32*, 429-492.

Everett, D. (2005). Biology and language: a consideration of alternatives, *Journal of Linguistics, 41*, 157-175.

Farrar, J. (1992). Negative evidence and grammatical morpheme acquisition. *Developmental Psychology, 28*, 90-98.

Fernald, A. (1991). Prosody in speech to children: Prelinguistic and linguistic functions. *Annals of Child Development, 8,* 43-80.

Fernald, A., & Hurtado, N. (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science, 9,* F33–F40.

Fiengo, R. (2006). Review of Chomsky's Minimalism by Pieter Seuren. *Mind, 115,* 469-472

Finch, S., & Chater, N. (1991). A hybrid approach to the automatic learning of linguistic categories. *Artificial Intelligence Society of Britain Quarterly, 78,* 16-24.

Finch, S., & Chater, N. (1994). Learning syntactic categories: a statistical approach. In M. Oaksford & G. Brown (Eds.), *Neurodynamics and Psychology* (pp. 294-321). London: Academic Press.

Fiser, J., & Aslin, R. (2001). Unsupervised statistical learning of higher-order spatial structures from visual scenes. *Psychological Science, 12,* 499-504.

Fitch, T., Hauser, M., & Chomsky, N. (2005). The evolution of the language faculty: Clarifications and implications. *Cognition 97,* 179–210.

Flage, D. & Bonnen, C. (1999). *Descartes and Method*. London: Routledge.

Fodor, J. (1975) The Language of Thought. Thomas Y. Crowell Co. Harvard University Press.

Fodor, J. (2000). *The Mind Doesn't Work That Way*. Cambridge: MIT Press.

Fodor, J. & Katz, J. (1964). *The structure of Language*. Prentice Hall.

Formigari, L. (2004). *A History of Language Philosophies.* Amsterdam: John Benjamins

Foti, V. (2000). Descartes' intellectual and corporeal memories. In: Gaukroger, S., Schuster, J., & Sutton, J. (Eds.) *Descartes' natural philosophy*. pp. 591 - 603. London: Routledge.

Fowler, C. (1999). *Descartes on the human soul.* Dordrecht: Kluwer Academic Publishers.

Freudenthal, D., Pine, J., & Gobet, F. (2010). Explaining quantitative variation in the rate of Optional Infinitive errors across languages: A comparison of MOSAIC and the Variational Learning Model. *Journal of Child Language, 37,* 643-669.

Gallistel, C. (2010). Learning Organs. In: Bricmont, J. & Frank, J. (Eds.) *Chomsky Notebook*. (pp. 193-202). New York: Columbia University Press.

Ganger, J. & Brent, M. (2004). Reexamining the Vocabulary Spurt. *Developmental Psychology, 40,* 621–632.

Gaukroger, S. (1992). Descartes's Early Doctrine of Clear and Distinct Ideas. *Journal of the History of Ideas, 53,* 585-602.

Gaukroger, S. (1995). *Descartes: An Intellectual Biography.* Oxford: Clarendon Press.

Gaukroger, S. (2000). The Foundational Role of Hydrostatics and Statics in Descartes' Natural Philosophy. In: Gaukroger, S., Schuster, J. & Sutton, J. (Eds.), *Descartes' Natural Philosophy*. (pp. 60 - 88). London: Routledge.

Gaukroger, S. (2002). *Descartes' System of Natural Philosophy*. Cambridge: Cambridge University Press.

Gaukroger, S. (2006). *The Blackwell guide to Descartes' Meditations*. Malden, MA: Blackwell.

Gelman, S., Chesnick, R. & Waxman, S. (2005) Mother–child conversations about pictures and objects: referring to categories and individuals. *Child Development 76 ,* 1129.

Gentner, D. & Goldin-Meadow S. (Eds). (2003). *Language in Mind*. Cambridge: MIT Press.

Gentner, T., Fenn, K., Margoliash, D., & Nusbaum, H. (2006) Recursive syntactic pattern learning by songbirds. *Nature, 440*, 1204-1207.

George, A. (1987). Review discussion on Chomsky's *Knowledge of Language. Mind and Language,* 2, 165 – 177.

Gewirth, A. (1967). Clearness and distinctness in Descartes. In: Doney, W. (Ed.) *Descartes - A collection of critical essays.* pp. 250 - 277. Garden City, NY: Anchor Books.

Gombay, A. (2007). *Descartes.* Oxford: Blackwell Publishing.

Gold, E. (1967). Language identification in the limit. *Information and Control 10*, 447–474

Goldin-Meadow, S., & Mylander, C. (1990). Beyond the input given: The child's role in the acquisition of language. *Language 66*, 323-355.

Goldstein, M., & West, M. (1999). Consistent responses of human mothers to prelinguistic infants: The effect of prelinguistic repertoire size. *Journal of Comparative Psychology, 113,* 52 - 58.

Goldstein, M., King, A., & West, J. (2003). Social interaction shapes babbling: Testing parallels between birdsong and speech. *Proceedings of the National Academy of Sciences, 100,* 8030 - 8035.

Goldwater, S., Griffiths, T. L. & Johnson, M. (2009). A Bayesian framework for word segmentation : Exploring the effects of context. *Cognition, 11*, 21–54.

Gómez, R., & Gerken, L. (1999). Artificial grammar learning by one-year-olds leads to specific and abstract knowledge. *Cognition, 70,* 109–135.

Gómez, R., & Gerken, L. (2000). Infant artificial language learning and language acquisition *Trends in Cognitive Sciences, 4,* 5.

Goodluck, H. (1991). *Language Acquisition: A Linguistic Introduction.* Oxford: Blackwell.

Goodman, N. (1965). *Fact, Fiction, and Forecast.* Indianapolis: Bobbs-Merrill.

Goodman, N. (1967). The Epistemological Argument. *Synthese, 17*, 23-28.

Goodman, N. (1968). The emperor's new ideas. In: Hook, S. (Ed.) *Language and Philosophy.* (pp. 138 -142). New York: New York University Press.

Gordon, P. (1990). Learnability and feedback: A commentary on Bohannon and Stanowicz. *Developmental Psychology, 26*, 215–218.

Graddol, D. (2004). The future of language. *Science, 303*, 1239–31.

Griffiths, P. (2002). What is Innateness? *The Monist, 85*, 70–85.

Grush, R. (2003). In Defense of Some 'Cartesian' Assumptions Concerning the Brain and Its Operation *Biology and Philosophy 18,* 53–93.

Guillaume, P. (1971). *Imitation in children.* Chicago: University of Chicago Press.

Gunderson, K. (1964). Descartes, La Mettrie, Language and Machines. *Philosophy 39,* 193-222.

Hall, D., & Graham, S. (1999). Lexical form class information guides word-to-object mapping in preschoolers. *Child Development*, 70, 78-91.

Hare, M., & Elman, J. (1995). Learning and morphological change. *Cognition, 56,* 61-98.

Harman, G. (1967). Psychological Aspects of the theory of syntax. *Journal of Philosophy 64,* 75-87.

Harman, G. (1968). Linguistic competence and empiricism. In: Hook, S. (Ed.) Language and Philosophy. (pp. 143 -151). New York: New York University Press.

Harman, G. (1974). Meaning and semantics. In: Munitz, M. & Unger, P. (Eds.), *Semantics and Philosophy*. (pp. 1-16). New York: New York University Press.

Harris, R. (1993). *The linguistic wars*. Oxford: Oxford University Press.

Harris, R. (1998). Review of *Noam Chomsky: A life of discent* by Robert Barsky, MIT Press. In: *Books in Canada 27*, 14-17.

Hatfield, G. (2008a). Mental Acts and Mechanistic Psychology in Descartes' Passions. In: Robertson, N., McOuat, G., & Vinci, T. (Eds.). *Descartes and the Modern*. pp. 49-71. Newcastle upon Tyne: Cambridge Scholars Publishing.

Hatfield, G. (2008b). Psychology in Philosophy: Historical Perspectives. In: Heinamaa, S. & Reuter, M. (Eds.), *Psychology and Philosophy*. Springer.

Hatfield, G. (2007). Did Descartes Have a Jamesian Theory of the Emotions?, *Philosophical Psychology 20*, 413–40.

Hatfield, G. (2003). *Descartes and the Meditations*. New York: Routledge.

Hatfield, G. (2000). Descartes' naturalism about the mental. In: Gaukroger, S., Schuster, J., & Sutton, J. (Eds.) *Descartes' natural philosophy.* pp. 630 - 658. London: Routledge.

Hauser, M. (1996). *The Evolution of Communication.* Cambridge. MA: MIT Press.

Hauser M., Newport E., & Aslin R. (2001). Segmentation of the speech stream in a nonhuman primate: Statistical learning in cotton-top tamarins. *Cognition, 78*, 53-64.

Hauser, M., Chomsky, N., & Fitch, W. (2002). The Faculty of Language: What Is It, Who Has It, and How Did It Evolve? *Science, 298*, 1569-1579.

Hepper, P. (2002). Prenatal Development. In A. Slater & M. Lewis (Eds.), *Introduction to Infant Development* (pp. 39-60). Oxford: Oxford University Press.

Hepper, P. & Shabidullah, B. (1994). Development of fetal hearing. Archives of Disease in Childhood-Fetal Neonatal Edition,71, F81-87.

Hirsh-Pasek, K., Treiman, R. & Schneidermann, M. (1984). Brown and Hanlon revisited: Mother's sensitivity to ungrammatical forms. *Journal of Child Language, 11*, 81-88.

Hirsh-Pasek, K., Tucker, M., & Golinkoff, R. (1996). Dynamic Systems Theory: Reinterpreting "Prosodic Bootstrapping" and its role in language acquisition. In: J. Morgan & K. Demuth (Eds.) *Signal to Syntax*. (pp. 449-466). Mahwah, NJ: Lawrence Erlbaum Associates.

Hirsh-Pasek, K., & Golinkoff, R. (1999). *The Origins of Grammar*. Cambridge, MA: MIT Press.

Hoffman, P. (1986). The Unity of Descartes's Man. *The Philosophical Review 95,* 339-370.

Hoffman, P. (2002). Descartes's Theory of Distinction. *Philosophy and Phenomenological Research, 64,* 57-78.

Hoffman, P. (2007). Descartes's Watch Analogy. *British Journal for the History of Philosophy, 15,* 561-567.

Hook, S. (1968). Empiricism, rationalism and innate ideas. In: Hook, S. (Ed.) Language and Philosophy. (pp. 160 -167). New York: New York University Press.

Hornstein, N. (2005). Chomsky and His Critics. *Croatian Journal of Philosophy 5,* 126-144.

Hornstein, N., & Lightfoot, D. (1981). *Explanation in linguistics: The logical problem of language acquisition*. London: Longman.

369

Hornstein, N. & Boeckx, C. (2009). Approaching Universals from below: I-Universals in light of a minimalist program for linguistic theory. In: Christiansen, M., Collins, C. & Edelman, S. (Eds.) *Language Universals*. (pp. 79 - 98). Oxford: Oxford University Press.

Hume, D. (1739/2000) *A Treatise of Human Nature.* Norton, D. & Norton, M. (Eds.) Oxford: Oxford University Press.

Hume, D. (1777/1958). An enquiry concerning human understanding. LaSalle, IL: Open Court Publishing.

Hymers, M. (2005). *Chomsky and the Poverty of Evidence*. Typescript. Dalhousie University.

Ingham, R. (1993). Critical influence on the acquisition of verb transitivity. In D. Messner & G. Turner (Eds.). Critical influence on child language acquisition and development. London: Maximillian Press.

Jablonka, E., & Lamb, M. (2005). *Evolution in Four Dimensions.* Cambridge, MA: MIT Press.

Jackendoff, R., & Pinker, S. (2005). The nature of the language faculty and its implications for evolution of language (reply to Fitch, Hauser, and Chomsky). *Cognition 97*, 211–225.

Jackendoff, R. (2006). Locating Meaning in the Mind (Where it Belongs) (excerpt from Foundations of Language) In: R. Stainton. (Ed.) Contemporary Debates in Cognitive Science. (pp. 219-236). Malden, MA: Blackwell.

Jackendoff, R. (2007). Language, Consciousness, Culture. Cambridge, MA: MIT Press.

Jackendoff, R. (2007). Linguistics in cognitive science: the state of the art. *The Linguistic Review* 24, 347-401.

Jackson, F. (1982). Epiphenomenal Qualia, *Philosophical Quarterly, 32,* 127-136.

Jaroz, G. (2010). Implicational markedness and frequency in constraint-based computational models of phonological learning *Journal of Child Language, 37,* 565-606.

Jay, T. (2003). *The Psychology of Language*. Upper Saddle River, NJ: Pearson Education.

Johansson, S. (2005). *Origins of Language, Constraints on Hypotheses*. Amsterdam: John Benjamins.

Johnson, E., & Jusczyk, P. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 1-20.

Johnson, K. (2004). Gold's theorem and cognitive science. *Philosophy of Science, 71*, 571–592.

Jolley, N. (1988). Leibniz and Malebranche on innate ideas. *The Philosophical Review, 97,* 71 - 91.

Joly, A. (1985). Cartesian or Condillacian Linguistics? *Topoi, 4*, 145-149.

Jusczyk, P. & Kemler Nelson, D. (1996) Syntactic units, prosody, and psychological reality during infancy. In: J. Morgan & K. Demuth (Eds.) *Signal to Syntax*. (pp. 389-408). Mahwah, NJ: Lawrence Erlbaum Associates.

Jusczyk, P., Houston, D., & Newsome, N. (1999). The beginnings of word segmentation in English-learning infants. *Cognitive Psychology*, *39,* 159-207.

Kasher, A. (1991). Pragmatics and Chomsky's Research Program. In: Kasher A (Ed). *The Chomskyan Turn*. (pp. 122-149). Oxford: Basil Blackwell.

Katz, J. (1971). *The underlying reality of language*. New York: Harper & Row.

Katz, J. (1981). *Language and other abstract objects*. Totowa, NJ: Rowman and Littlefield.

Katz, J. (1985). An outline of Platonist grammar. In: Katz, J. (Ed.) *The philosophy of linguistics*. (pp. 172-203). Oxford: Oxford University Press.

Katz, J. (1986). *Cogitations*. New York: Oxford University Press.

Katz, J. (1990). *The metaphysics of meaning*. Cambridge, MA: MIT Press.

Katz, J. & Postal, P. (1991). Realism vs. Conceptualism in Linguistics. *Linguistics and Philosophy 14*, 515-554.

Katz, J. (1996). The unfinished Chomskyan revolution. *Mind and Language 11*, 270-294.

Katz, J. (1998). *Realistic Rationalism*. Cambridge, MA: MIT Press.

Katz, J. (2004). *Sense, reference, and philosophy*. New York: Oxford University Press.

Kelly, M. (1992). Using sound to solve syntactic problems: The role of phonology in grammatical category assignments. *Psychological Review, 99,* 349-364.

Kenny, A. (1967). Descartes on Ideas. In: Doney, W. (Ed.) *Descartes - A collection of critical essays.* pp. 227 - 249. Garden City, NY: Anchor Books.

Kenny, A. (1968). *Descartes: A Study of His Philosophy.* New York: Random House.

Kenney, A. (1972). *Descartes.* Oxford: Oxford University Press.

Kimball, J. (1973). The Formal Theory of Grammar. Englewood Cliffs, NJ: Prentice Hall.

Kirby, S. & Christiansen, M. (2003). From language learning to language evolution. In M. Christiansen & Kirby, S. (Eds.), *Language Evolution.* (pp. 272-294). Oxford: Oxford University Press.

Kirby, S., & Hurford, J. (1997) Learning, Culture, and Evolution in the Origin of Linguistic Constraints. In P. Husbands & I. Harvey. (Eds). *Fourth European Conference on Artificial Life* (pp.493-502). Cambridge. MA: MIT Press.

Koopman, H. (2000). *The syntax of specifiers and heads*. London: Routledge.

Lakoff, G. & Ross, J. (1976). Is Deep Structure Necessary. In: McCawley, J. (Ed.) *Notes from the Linguistic Underground, Syntax and Semantics* 7, New York, Academic Press.

Lakoff, R. (1969). Review of H. Brekle (Ed.) Grammaire generale et raisonnee. *Language, 45,* 343-364.

Langendoen, T. & Postal, P. (1984). *The vastness of natural language*. Oxford: Basil Blackwell.

Lasky, R. & Williams, A. (2005). The development of the auditory system from conception to term. *Neoreviews, 6*, 141-152.

Lecanuet, J. & Schaal, B. (1996). Fetal sensory competencies. *European Journal of Obstestrics, Gynecology and Reproductive Biology, 68*, 1-23.

Legate, J. & Yang, C. (2002). Empirical Re-Assessment of Stimulus Poverty Arguments. *Linguistic Review 19*, 151-162.

Leiber, J. (1988). Cartesian Linguistics? In: The Chomskyan Turn: Generative Linguistics, Mathematics, Philosophy and Psychology International Workshop, Tel Aviv, *Philosophia, 1988, 309-46.*

Lenneberg, E. (1967). *Biological Foundations of Language*. New York: Wiley.

Leontev, A. (1975). *Psycholinguistische Einheiten und die Erzeugung sprachlicher Aeusserungen.* Berlin: Max Hueber Verlag

Levine, R. & Postal. P. (2004). A Corrupted Linguistics. In P. Collier & D. Horowitz, (Eds.). *The Anti-Chomsky Reader*. (p.p. 203-32). San Francisco: Encounter Books.

Lewis, D. (1975). Languages and Language. In: Gunderstone, K. (Ed.) Minnesota Studies in the Philosophy of Science Volume VII *Language Mind, and Knowledge*. pp. Minneapolis: University of Minnesota Press.

Lidz, J., Waxman, S., & Freedman, J. (2003). What infants know about syntax but couldn't have learned: Evidence for syntactic structure at 18-months. *Cognition, 89*, B65–B73.

Lidz, J. & Waxman, S. (2004). Reaffirming the poverty of the stimulus argument: a reply to the replies. *Cognition, 93,* 157-165.

Lieberman, P. (2007). The Evolution of Human Speech: Its Anatomical and Neural Bases. *Current Anthropology*, 48, 39-66.

Lieven, E. (1994). Crosslinguistic and crosscultural aspects of language addressed to children. In Gallaway, C., & Richards, B. (Eds.). *Input and interaction in language acquisition*. (pp. 56-73). Cambridge: Cambridge University Press.

Lieven, E., Behrens, H., Speares, J., & Tomasello, M. (2003). Early syntactic creativity: a usage-based approach. *Journal of Child Language, 30,* 333-370.

Lightfoot, D. (1989). The child's trigger experience: Degree-0 learnability *Behavioral & Brain Sciences, 12*, 321-334.

Lightfoot, D. (1991). *How to set Parameters: Arguments from Language Change.* Cambridge, MA: MIT Press.

David Lightfoot. (1999). The development of language: Acquisition, change and evolution. Oxford: Blackwell,

Lightfoot, D. (2002). Myths and the prehistory of grammars. *Journal of Linguistics, 38*, 113-136.

Lightfoot, D. (2005). Plato's Problem, UG, and the Language Organ. In: McGilvray, J. (Ed.) *The Cambridge Companion to Chomsky* (pp. 42-59). Cambridge: Cambridge University Press.

Lindeboom, G. (1978). *Descartes and Medicine*. Amsterdam: Rodopi.

Losonsky, M. (2006). *Linguistic Turns in Modern Philosophy*. Cambridge: Cambridge University Press.

Lyons, J. (1970). *Chomsky*. London: Fontana Modern Masters.

MacDonald Ross, G. (1988). Hobbes and Descartes on the relation between language and consciousness. *Synthese, 75,* 217-229.

MacWhinney, B., & Snow, C. (1985). The child language data exchange system. *Journal of Child Language, 12*, 271-296.

MacWhinney, B., & Snow, C. (1990). The child language data exchange system: An update. *Journal of Child Language, 17*, 457-472.

MacWhinney, B. (1995). *The CHILDES project: tools for analyzing talk.* Hillsdale, NY: Lawrence Erlbaum Associates.

MacWhinney, B. (2000) *The CHILDES Project: Tools for Analyzing Talk.* Mahwah, NJ: Lawrence Erlbaum Associates.

MacWhinney, B. (2001). From CHILDES to TalkBank. In M. Almgren, A. Barreña, M. Ezeizaberrena, I. Idiazabal & B. MacWhinney (Eds.), *Research on Child Language Acquisition*. (pp. 17-34). Somerville, MA: Cascadilla.

MacWhinney, B. (2004a) A Unified Model of Language Acquisition. In J. Kroll & A. De Groot (Eds.) *Handbook of bilingualism: Psycholinguistic approaches.* Oxford: Oxford University Press.

MacWhinney, B. (2004b). A multiple process solution to the logical problem of language acquisition. *Journal of Child Language 31*, 883–914.

MacWhinney, B. (2008). Enriching CHILDES for morphosyntactic analysis. In: Behrens, H. (Ed.) Corpora in Language Acquisition Research. (pp. 165 - 197). Amsterdam: John Benjamins.

MacWhinney, B. (2010). Computational models of child language learning: an introduction. *Journal of Child Language, 37,* 477-485.

Malcolm, N. (1968). *Rene Descartes "Discourse on Method and the Meditations"*. London: Penguin.

Malcolm, N. (1993). *Wittgenstein: A Religious Point Of View?* Ithaca: Cornell University Press.

Mameli, M., & Bateson, P. (2006). Innateness and the sciences. *Biology and Philosophy, 21,* 155–188.

Maratsos, M., Gudeman, R., Gerard-Ngo, P., & DeHart, G. (1987). A study in novel word learning: the productivity of the causative. In B. Mac Whinney (Ed.). *Mechanisms of language acquisition.* Hillsdale, NJ: Erlbaum.

Marcus, G. (1993). Negative evidence in language acquisition. *Cognition,* 46, 53-85.

Marcus, G. (1998). Rethinking eliminative connectionism. *Cognitive Psychology, 37*, 243-282.

Marcus, G. (1999). Language acquisition in the absence of explicit negative evidence: Can simple recurrent networks obviate the need for domain-specific learning devices*? Cognition, 73,* 293-296.

Marcus, G. (2001). Plasticity and nativism: Towards a resolution of an apparent paradox In: Wermter , S., AustinJ., & Willshaw, D. (Eds.) *Emergent neural computational architectures based on neuroscience.* (pp. 368-382.) New York: Springer-Verlag.

Marcus, G. (2006). Cognitive architecture and decent with modification. *Cognition, 101*, 443-465.

Marcus, G. & Brent, I. (2003). Are There Limits to Statistical Learning? *Science, 300*, 53-54.

Marion, J. (2008). Descartes and the horizon of finitude. In: Robertson, N., McOuat, G., & Vinci, T. (Eds.). *Descartes and the Modern*. pp. 196 - 215. Newcastle upon Tyne: Cambridge Scholars Publishing.

Markova, I. (1982). *Paradigms, Thought, and Language.* New York: John Wiley.

Markson, L., Diesendruck, G., & Bloom, P. (2008). The shape of thought. *Developmental Science, 11,* 204-208.

Marles, M. (1978). Doubt, reason and Cartesian therapy. In: Hooker, M. (Ed.) *Descartes: Critical and Interpretative Essays.* pp. 89 - 113. Baltimore: Johns Hopkins University Press.

Martin, R. (1986). *The Meaning of Language*. Cambridge. MA: MIT Press.

Martin, R. (2005). *Philosophical Conversations*. Peterborough: Broadview Press.

Matthews, R. (2005). The case for linguistic nativism. In. Stainton, R. (Ed.) *Contemporary debates in cognitive science*. (pp. 81 - 96). Malden, MA: Blackwell Publishing.

Mattys, S., Jusczyk, P., Luce, P., & Morgan, J. (1999). Word segmentation in infants: How phonotactics and prosody combine. *Cognitive Psychology, 38,* 465-494.

Mattys, S., & Jusczyk, P. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78,* 91-121.

Maye, J., Werker, J., & Gerken, L. (2002). Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition, 82,* B101–B111.

McCawley, J. (1976). *Grammar and meaning*. New York: Academic Press.

McClelland, J., McNaughton, B., & O'Reilly, R. (1995). Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychological Review, 102,* 419-457.

McDermott, D. (2001). *Mind and Mechanism*. Cambridge, MA: MIT Press.

McGilvray, J. (1999). *Chomsky: Language, mind and politics.* Cambridge, MA: Polity Press.

McGilvray, J. (2002). Introduction for cybereditions. In: Chomsky, N. *Cartesian Linguistics. A chapter in the history of Rationalist Thought.* Christchurch, NZ: Cybereditions.

McGilvray, J. (2005). The Cambridge companion to Chomsky. Cambridge: Cambridge University Press.

McGilvray, J. (2006). On the innateness of language. In. Stainton, R. (Ed.) *Contemporary debates in cognitive science*. (pp. 97 - 112). Malden, MA: Blackwell Publishing.

McGilvray, J. (2009). Introduction to the third edition. In: Chomsky, N. *Cartesian Linguistics. A chapter in the history of Rationalist Thought.* Cambridge: Cambridge University Press.

McGraw, T. (1992). Unravelling innate ideas. *History of Philosophy Quarterly, 9,* 307 - 317.

McMurray, B. (2007) Defusing the childhood vocabulary explosion. *Science, 317,* 631.

McMurray, B., & Aslin, R. (2005) Infants are sensitive to within-category variation in speech perception. *Cognition, 95,2,* B15-B26.

McNeill, D. (2005). *Gesture and Thought.* Chicago: University of Chicago Press.

McRae, R. (1965). 'Idea' as a philosophical term n the seventeenth century. *Journal of the History of Ideas, 26,* 175 - 190.

McRae, R. (1972). Innate Ideas. In: Buttler, R. (Ed.) *Cartesian Studies.* pp. 32-54. Oxford: Blackwell.

Mehta, V. (1971). *John Is Easy to Please*. London: Secker & Warburg.

Miel, J. (1969). Pascal, Port-Royal, and Cartesian Linguistics. *Journal of the History of Ideas, 30,* 261-271.

Miikkulainen, R. & Mayberry, M.(1999). Disambiguation and grammar as emergent soft constraints. In: B. MacWhinney (Ed.). The Emeregence of Language. (pp. 153-176). Mahwah, NJ: Lawrence Erlbaum Associates.

Mintz, T. (2002) Category induction from distributional cues in an artificial language. *Memory & Cognition, 30,* 678-686.

Mintz, T. (2006). Frequent Frames: Simple Co-occurrence constructions and their links to linguistic structure. In: E. Clark & B. Kelly (Eds.) Constructions in Acquisition. Stanford: CSLI Publications.

Moerk, E. (1992). *A First Language Taught and Learned.* Baltimore: Brooks Publishing.

Moerk, E. & Moerk, C. (1979). Quotations, Imitations, and Generalizations. Factual and Methodological Analyses. *International Journal of Behavioral Development, 2*, 43-72.

Monaghan, P., Chater, N., & Christiansen, H. (2005). The differential contribution of phonological and distributional cues in grammatical categorization. *Cognition, 96,* 143-182.

Monaghan, P. & Christiansen, M. (2008). Integration of multiple probabilistic cues in syntax acquisition. In H. Behrens (Ed.), *Trends in corpus research: Finding structure in data.* (pp. 139-163). Amsterdam: John Benjamins.

Monaghan, P., & Christiansen, M. (2010). Words in puddles of sound: modelling psycholinguistic effects in speech segmentation. *Journal of Child Language, 37,* 545-564.

Moon, C.M. & Fifer, W.P. (2000). The fetus: Evidence of transnatal auditory learning. *Journal of Perinatology, 20*, S37–S44.

Morgan, J., Meier, R., & Newport, E. (1987). Structural packaging in the input to language learning: Contributions of prosodic and morphological marking of phrases to the acquisition of language. *Cognitive Psychology, 19*, 498-550.

Morris, J. (1971). *Descartes Dictionary*. New York: Philosophical Library.

Morris, K. (2000). Betes-machines. In: Gaukroger, S., Schuster, J., & Sutton, J. (Eds.) *Descartes' natural philosophy*. pp. 401 - 419. London: Routledge.

Mukherji, N. (2010). Review of Jean Bricmont and Julie Franck (Eds.), Chomsky Notebook, Columbia University Press. In: *Notre Dame Philosophical Reviews*.

Murray, S. (2010). "Scientific revolutions" and other kinds of regime change. In: Kibbee, D. (Ed.) Chomskyan (R)evolutions. (pp. 75 – 101). Amsterdam: John Benjamins.

Musolino, J., Crain, S., & Thornton, R. (2000). Navigating Negative Semantic Space. *Linguistics, 38,* 1-32.

Musgrave, A. (1993). *Common Sense, Science, and Scepticism.* Cambridge: Cambridge University Press.

Nadler, S. (2006). The doctrine of ideas. In: Gaukroger, S. (Ed.) *The Blackwell Guide to Descartes' Meditations.* pp. 86-103. Oxford: Blackwell Publishing.

Nadler, S. (1989). *Arnauld and the Cartesian Philosophy of Ideas*. Princeton: Princeton University Press.

Nazzi, T., Bertoncini, J., & Mehler, J. (1998). Language discrimination by newborns: toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human Perception and Performance, 24*, 756-66.

Nazzi, T., Jusczyk, P., & Johnson, E. (2000a). Language discrimination by English-learning 5-month-olds: Effect of rhythm and familiarity. *Journal of Memory and Language, 43*, 1-19.

Nazzi. T., Kemler Nelson, D., Jusczyk, P., & Jusczyk, A. (2000b). Six-month-olds' detection of clauses embedded in continuous speech: Effects of prosodic wellformedness. *Infancy, 1,* 123-147.

Nelson, K. (1987). Some observations from the perspective of the rare event cognitive comparison theory of syntax acquisition. In K. Nelson (Ed.), *Children's language* (pp. 289–331). Hillsdale, NJ: Erlbaum.

Nelson, A. (1997). Descartes' ontology of thought. *Topoi 16,* 163-178.

Newman, R. (2005). The cocktail party effect in infants revisited: Listening to one's name in noise. *Developmental Psychology 41,* 352–62.

Newman, R., Bernstein Ratner, N., Jusczyk, A., Jusczyk, P. & Dow, K. (2006). Infants early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental Psychology 42,* 643–55.

Newmeyer, F. (2003). Review article: Chomsky, On nature and language; Anderson and Lightfoot, The language organ; Bichakjian, Language in a Darwinian perspective. *Language*, 79, 583 – 599.

Newmeyer, F. (2008). Linguistic minimalism: Origins, concepts, methods, and aims. *Language, 84,* 387-395.

Newport, E., Gleitman, H., &. Gleitman, L. (1977). Mother, I'd rather do it myself: Some effects and non-effects of maternal speech style. In C. Snow & C. Ferguson (Eds.), *Talking to Children: Language Input and Acquisition.* (pp. 109-150). Cambridge: Cambridge University Press.

Nolan, L. (1997). The Ontological Status of Cartesian Natures. *Pacific Philosophical Quarterly, 78*, 169-94.

Norman, M. (1967). Descartes' proof that his essence is thinking. In: Doney, W. (Ed.) *Descartes - A collection of critical essays.* pp. 312 - 337. Garden City, NY: Anchor Books.

Normore, C. (1986). Meaning and Objective Being: Descartes and His Sources. In: Oksenberg Rorty, A (Ed.) *Essays on Descartes' Meditations.* pp. 223-41. Berkeley: University of California Press.

Onnis, L., Christiansen, M. & Chater, N. (2009). Connectionist models of language processing. In: L.R. Squire (Ed.) *New encyclopedia of neuroscience. Vol. 3* (pp. 83-90). Oxford, U.K.: Elsevier.

Onnis, L., Roberts, M. & Chater, N. (2002). Simplicity: A cure for overregularizations in language acquisition? In W. Gray & C. Schunn (Eds.), *Proceedings of the 24th Annual Conference of the Cognitive Science Society.* (pp. 720–25). Austin, TX : Cognitive Science Society.

Orlov, T., Yakolev, V., Hochstein, S., & Zohary, E. (2000). Macaque monkeys categorize images by their ordinal number. *Nataure 404,* 77-80.

Otero, C. (1988). *Noam Chomsky, Language and Politics*. Montreal: Black Rose Books.

Pacton, S., Perruchet, P., Fayol, M., & Cleeremans, A. (2001). Implicit learning in real world context: The case of orthographic regularities. *Journal of Experimental Psychology: General, 130,* 401-426.

Percival, K. (1972). On the non-existence of Cartesian linguistics. In: Buttler, R. (Ed.). *Cartesian Studies.* pp. 137 - 145. Oxford: Basil Blackwell.

Perfors, A., Tenenbaum, J., & Wonnacott, E. (2010). Variability, negative evidence, and the acquisition of verb argument constructions. *Journal of Child Language, 37,* 607-642.

Perruchet, P., & Vinter, A. (2002). The Self-Organizing Consciousness. *Behavioral and Brain Sciences, 25,* 297- 330.

Perruchet, P., & Rey, A. (2005). Does the mastery of center-embedded linguistic structures distinguish humans from nonhuman primates? *Psychonomic Bulletin and Review, 12,* 307-313.

Perruchet, P., & Pacton, S. (2006). Implicit learning and statistical learning : Two approaches, one phenomenon. *Trends in cognitive science, 10,* 233-238.

Perruchet, P & Tillmann, B. (in press). Exploiting multiple sources of information in learning an artificial language: Human data and modeling. *Cognitive Science*.

Pfeifer, R. & Bongard, J. (2005). *How the Body Shapes the Way we Think.* Cambridge, MA: MIT Press.

Pietroski, P. & Crain, S. (2005). Innate Ideas. In: J. McGilvray (Ed.). The Cambridge Companion to Chomsky. Cambridge: Cambridge University Press.

Pinker, S. (1984). *Language learnability and language development*. Cambridge, MA : Harvard University Press.

Pinker, S. (1989). *Learnability and Cognition*. Cambridge, MA: MIT Press.

Pinker, S. (1994). *The language instinct*. New York: HarperCollins

Pinker, S. (1998). Words and rules. *Lingua, 106,* 219-242.

Pinker, S. (2004) Clarifying the logical problem of language acquisition. *Journal of Child Language, 31*, 949-953.

Plunkett, K., & Marchman, V. (1993). From rote learning to system building: Acquiring verb morphology in children and connectionist nets. *Cognition, 48,* 21-69.

Postal, P. (1974a). On certain ambiguities. *Linguistic Inquiry, 5*, 367-424.

Postal, P. (1974b). *On raising*. Cambridge, MA: MIT Press.

Postal, P. (2003a). Remarks on the foundations of linguistics. *The philosophical forum, 34*, 233-251.

Postal, P. (2003b). (Virtually) Conceptually Necessary. *Journal of Linguistics 39*, 1-22.

Postal, P. (2004). *Skeptical linguistic essays*. New York: Oxford University Press.

Postal, P. (2009). The Incoherence of Chomsky's 'Biolinguistic' Ontology. *Biolinguistics 3,* 104–123.

Pullum, G. (1972). Grammars, competence and linguistic intuitions. *Linguistische Berichte, 21*, 55-64

Pullum, G. (1991). *The Great Eskimo Vocabulary Hoax*. Chicago: University of Chicago Press.

Pullum, G. (1996). Learnability, hyperlearning, and the poverty of the stimulus. In: Jan Johnson, Matthew L. Juge, and Jeri L. Moxley (eds.), *Proceedings of the 22nd Annual Meeting: General Session and Parasession on the Role of Learnability in Grammatical Theory*, 498-513. Berkeley Linguistics Society, Berkeley, California.

Pullum G., & Scholz B., (2002). Empirical assessment of stimulus poverty arguments. *The Linguistic Review, 19,* 9-50.

Pullum, G. (2007). Ungrammaticality, rarity, and corpus use. *Corpus Linguistics and Linguistic Theory 3*, 33-47.

Putnam, H. (1967). The 'Innateness Hypothesis' and Explanatory Models in Linguistics. *Synthese, 17*, 12-22.

Putnam, H. (1988). *Representation and Reality*. Cambridge, MA: MIT Press

Quine, W. (1968). Linguistics and Philosophy. In: Hook, S. (Ed.) *Language and Philosophy*. (pp. 95 - 98). New York: New York University Press.

Ramsey, W. & Stich, S. (1991). Connectionism and three levels of nativism. In: Ramsey, W., Stich, S., & Rummelhart (Eds.).*Philosophy and Connectionist Theory*. (pp. 287-310). Hillsdale, NJ: Lawrence Erlbaum Associates.

Ramus, F., Hauser, M., Miller, C., Morris, D., & Mehler, J. (2000) Language discrimination by human newborns and by cotton-top tamarin monkeys. *Science, 288*, 340-351.

Reali, F., & Christiansen, M. (2005). Uncovering the richness of the stimulus: Structural dependence and indirect statistical evidence. *Cognitive Science*, 29, 1007-1028.

Redington, M., & Chater, N. (1997). Probabilistic and distributional approaches to language acquisition. *Trends in Cognitive Sciences, 1*, 273-281.

Redington, M., Chater, N. & Finch, S. (1998). Distributional information: A powerful cue for acquiring syntactic categories. *Cognitive Science, 22,* 425-469.

Ree, J. (1974). *Descartes*. London: Allen Lane.

Riskin, J. (2003). The defecating duck, or, the ambiguous origins of artificial life. *Critical Inquiry, 29*, 599-633.

Rohde, D., & Plaut, D. (2003). Less is less in language acquisition. In P. Quinlan (Ed.), *Connectionist modelling of cognitive development* (pp. 189-231). Hove, UK: Psychology Press.

Rohde, D., & Plaut, D.C. (2003). Connectionist models of language processing. *Cognitive Studies, Japan, 10*, 10-28.

Rodis-Lewis, G. (1978). Limitations of the mechanical model in the Cartesian conception of the organism. In: Hooker, M. (Ed.) *Descartes: Critical and Interpretative Essays.* pp. 152 - 170. Baltimore: John Hopkins University Press.

Roy, D. (2007). *"http://www.media.mit.edu/cogmac/projects/hsp.html"*

Rozemond, M. (2006). The nature of the mind. In: Gaukroger, S. (Ed.) *The Blackwell Guide to Descartes' Meditations.* pp. 48-66. Oxford: Blackwell Publishing.

Rozemond, M. (1998). *Descartes's Dualism.* Cambridge MA: Harvard University Press.

Rubino, R. & Pine, J. (1998). Subject-verb agreement in Brazilian Portugese: What low error rates hide. *Journal of Child Language, 25*, 35-60.

Russell, J. (2004). *What is language development?.* Oxford: Oxford University Press.

Rytting, C., Brew, C., & Fosler-Lussier, E. (2010). Segmenting words from natural speech: subsegmental variation in segmental cues. *Journal of Child Language, 37,* 513-543.

Sagae, K., MacWhinney, B., & Lavie, A. (2004). Automatic parsing of parent-child interactions. Behavior Research Methods, Instruments, and Computers, 36, 113-126.

Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Wintner, S. (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language, 37,* 705-729.

Saffran, J., Aslin, R., & Newport, E. (1996). Statistical learning by 8-month old infants. *Science, 274,* 1926-1928.

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

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

Saffran, J. (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science, 12*, 110-114.

Saffran, J., Werker, J., & Werner, L. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (Eds.). *Handbook of Child Development.* (pp.58-108). New York: Wiley.

Salmon, V. (1969). Review of N. Chomsky Cartesian Linguistics. *Journal of Linguistics, 5*, 165-167.

Sampson, G. (1975). *The form of language*. London: Weidenfeld and Nicholson.

Sampson, G. (1980). *Schools of Linguistics: Competition and Evolution.* London*:* Hutchinson.

Sampson, G. (1989). Language acquisition: growth or learning? *Philosophical Papers 18,* 203 - 240.

Sampson, G. (1997). *Educating Eve: The 'Language Instinct'* Debate London: Continuum.

Sampson, G. (2001). *Empirical Linguistics*. London: Continuum.

Sampson, G. (2002). Exploring the richness of the stimulus. *The Linguistic Review, 19,* 73-104.

Sampson, G. (2005). Quantifying the shift towards empirical methods. *International Journal of Corpus Linguistics, 10*, 15–36.

Saxton, M. (1997). The Contrast Theory of negative input. *Journal of Child Language, 24*, 139-161.

Saxton, M. (2000). Negative evidence and negative feedback: immediate effects on the grammaticality of child speech, *First Language* 20, 221-252.

Saxton, M., Kulcsar, B., Marshall, G., & Rupra, M. (1998). The longer-term effects of corrective input: An experimental approach. *Journal of Child Language, 25*, 701–721.

Saxton, M., Houston-Price, C. & Dawson, N. (2005). The prompt hypothesis: clarification requests as corrective input for grammatical errors, *Applied Psycholinguistics* 26/3, 393-414.

Scholz, B. & Pullum, G., (2002). Searching for arguments to support linguistic nativism. *The Linguistic Review, 19,* 185-223.

Scholz, B. & Pullum, G. (2006). Irrational nativist exuberance. In. Stainton, R. (Ed.) *Contemporary debates in cognitive science*. (pp. 59 - 80). Malden, MA: Blackwell Publishing.

Seidenberg, M., & MacDonald, M. (1999) A probabilistic constraints approach to language acquisition and processing. *Cognitive Science, 23,* 569-588.

Seidenberg, M., MacDonald, M., & Saffran, J. (2002). Does grammar start where statistics stop? *Science, 298*, 553-554.

Sepper, D. (1996). *Descartes' Imagination: Proportion, Images, and the Activity of Thinking*. Berkeley: University of California Press.

Seris, J.-P. (1993). Language and machine in the philosophy of Descartes. In: Voss, S. (Ed.). *Essays in the Philosophy of René Descartes.* pp. 177 - 192. Oxford: Oxford University Press.

Seuren, P. (1968/2010). *Operator and Nucleus*. Cambridge: Cambridge University Press.

Seuren, P. (1998). *Western Linguistics*. Oxford: Blackwell Publishing.

Seuren, P. (2004). *Chomsky's Minimalism*. New York: Oxford University Press.

Shi, R., Morgan, J., & Allopena, P. (1998). Phonological and acoustic bases for earliest grammatical category assignment: A cross-linguistic perspective. *Journal of Child Language, 25,* 169-201.

Shi, R., Werker, J., & Morgan, J. (1999). Newborn infants' sensitivity to perceptual cues to lexical and grammatical words. *Cognition, 72*, B11–B21.

Shi, R., & Werker, J. (2003). The basis of preference for lexical words in 6-month-old infants. *Developmental Science, 6*, 484-488.

Skinner, B. (1957). *Verbal Behavior.* New York: Appleton Century Crofts.

Skirry, J, (2005). *Descartes and the Metaphysics of Human Nature*. London: Thoemmes-Continuum Press.

Slater, A. & Lewis, M. (2002). *Introduction to Infant Development.* Oxford: Oxford University Press.

Schmaltz, T. (1997). Descartes on innate ideas, sensation and scholasticism: the response to Regius. In: Steward, M. (Ed.) *Studies in Seventeenth-Century Philosophy.* pp. 33-73. Oxford: Clarendon Press.

Schmaltz, T. (2002). *Radical Cartesianism.* Cambridge: Cambridge University Press.

Schmaltz, T. (2006). Seventeenth-century responses to the Meditations. In: Gaukroger, S. (Ed.) *The Blackwell Guide to Descartes' Meditations.* pp. 193 - 203. Oxford: Blackwell Publishing.

Schmaltz, T. (2008). *Descartes on Causation.* Oxford: Oxford University Press.

Smeeton, N., Williams, A., Hodges, N., & Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. *Journal of Experimental Psychology: Applied, 11*, 98-110.

Smith, N. (1999) *Chomsky. Ideas and Ideals*. Cambridge: Cambridge University Press.

Smith, N. (2000). Foreword. In: Chomsky, N. *New horizons in the study of language and mind.* (pp. vi-xvi). Cambridge: Cambridge University Press.

Smith, N. (2005). Chomsky's science of language. In. McGilvray, J. (Ed.). *The Cambridge Companion to Chomsky*. (pp. 21-41). Cambridge: Cambridge University Press.

Sorell, T. (1987). *Descartes.* Oxford: Oxford University Press.

Sorell, T. (2005). *Descartes reinvented.* Cambridge: Cambridge University Press.

Speidel, G. & Nelson, K. (1989). *The many faces of imitation in language learning.* New York: Springer-Verlag.

Stainton, R. (1996). *Philosophical perspectives on language.* Peterborough, ON: Broadview.

Stainton, R. (2006). *Words and Thought*. Oxford: Clarendon Press.

Stich, S. (1975). Introduction: The Idea of Innateness. In: Stich, S. (Ed.) *Innate ideas*. pp. 1-22. Berkeley: University of California Press.

Sutton, J. (2000). The body and the brain. In: Gaukroger, S., Schuster, J., & Sutton, J. (Eds.) *Descartes' natural philosophy.* pp. 697 - 722. London: Routledge.

Tallerman, M. (Ed.) (2005). Language Origins: Perspectives on Evolution. Oxford: Oxford University Press.

Terrace, H. (2001). Chunking and serially organized behavior in pigeons, monkeys and humans. In R. Cook (Ed.). *Avian visual cognition* [On-line]. Available: www.pigeon.psy.tufts.edu/avc/terrace/

Thorndike, E. (1932). *The Fundamentals of Learning.* New York: Teachers College Press.

Tlumak, J. (1978). Certainty and Cartesian method. In: Hooker, M. (Ed.) *Descartes: Critical and Interpretative Essays.* pp. 40 - 73. Baltimore: John Hopkins University Press.

Tomasello, M. (1992a). The Social Bases of Language Acquisition. *Social Development, 1*, 67-87.

Tomasello, M. (1992 b). *First verbs: A case study in early grammatical development.* Cambridge: Cambridge University Press.

Tomasello, M. (2000). Do young children have adult syntactic competence? *Cognition, 74*, 209-253.

Tomasello, M. (2003). *Constructing a language: a usage-based theory of language acquisition.* Cambridge, MA: Harvard University Press.

Tomasello, M., Akhtar, N., Dodson, K., & Rekau, L. (1997). Differential productivity in young children's use of nouns and verbs. *Journal of Child Language, 24*, 373-87.

Tomasello, M. & Stahl, D. (2004). Sampling children's spontaneous speech: How much is enough? *Journal of Child Language, 31*, 101-121.

Tomasello, M. (2004). What kind of evidence could refute the UG hypothesis? *Studies in Language, 28,* 642-44.

Tomasello, M. (2005). Beyond formalities: The case of language acquisition *The Linguistic Review 22,* 183–197.

Tomasello, M. (2009). *Origins of communication*. Cambridge, MA: MIT Press.

Trabant, J. (2003). *Mithidates im Paradies: Kleine Geschichte des Sprachdenkens.* Muenchen: C. H. Beck Verlag.

Turing, A., 1950, Computing machinery and intelligence, *Mind, 50,* 433-460.

Valian, V. (1999). Input and language acquisition. In W. Ritchie & T. Bhatia (Eds.), *Handbook of child language acquisition*. (pp. 497-530). New York: Academic Press.

Vallabha, G., McClelland, J., Pons, F., Werker, J., & Amano, S. (2007). Unsupervised learning of vowel categories from infant-directed speech. *PNAS, 104, 33,* 13273-13278.

van de Pitte, F. (1985). Descartes' innate ideas. *Kant Studien, 76,* 363-384.

van de Weijer, J. (1999). Language Input for Word Discovery. Doctoral dissertation, Njimegen.

Van Rij, J., Van Rij, H. & Hendriks, P. (2010). Cognitive architectures and language acquisition: A case study in pronoun comprehension. *Journal of Child Language, 37,* 731-766.

Vinci, T. (1998). *Cartesian Truth.* Oxford: Oxford University Press.

Vinci, T. (2008). Descartes and Contemporary Epistemology. In: Robertson, N., McOuat, G., & Vinci, T. (Eds.). *Descartes and the Modern.* pp. 254-283. Newcastle upon Tyne: Cambridge Scholars Publishing.

Wagner, K (1985). How much do children say in a day? *Journal of Child Language, 12,* 475-487.

Waterfall, H., Sandank, B., Onnis, L., & Edelman, S. (2010). An empirical generative framework for computational modeling of language acquisition. *Journal of Child Language, 37,* 671-703.

Wells, N. (1990). Objective reality of ideas in Descartes, Caterus and Suarez. *Journal of the History of Philosophy, 28,* 3 - 61.

Wells, R. (1968). Innate Knowledge. In: Hook, S. (Ed.) *Language and Philosophy.* pp. 99-119. New York: New York University Press.

Werker, J., & Tees, R. (1984). Cross language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development*, 7, 49-63.

Werker, J., & Tees, R. (1999). Influences on infant speech processing: toward a new synthesis *Annual Review of Psychology, 50,* 509-535.

Wexler, K. & Culicover, P. (1980). *Formal Principles of Language Acquisition*. Cambridge: MIT Press.

Wildgen, W. (2004). *The Evolution of Human Language*. Amsterdam: John Benjamins Publishing Company.

Windelband, W. (1958). *A History of Philosophy - Volume II: Renaissance, Enlightenment, and Modern*. New York: Harper & Brothers.

Williams, B. (2005). *Descartes: the project of pure inquiry.* New York: Routledge.

Wilson, C. (1982). Sensation and Explanation: Descartes on Conscious Awareness. *Nature and System, 4,* 175-199.

Wilson, C. (2003). *Descartes' Meditations: An Introduction*. Cambridge: Cambridge University Press

Wilson, M. (1995). Animal Ideas. *Proceedings and Addresses of the American Philosophical Association, 69,* 7-25.

Wilson, M, (1978). *Descartes*. London: Routledge and Kegan Paul.

Winston, M. (2002). *On Chomsky* London: Wadsworth Philosophers Series

Wittgenstein, L. (1953). *Philosophical Investigations.* G. Anscombe & R. Rhees (Eds.), G. Anscombe (Trans.), Oxford: Blackwell.

Wonnacott, E., Newport, E. & Tanenhaus, M. (2008). Acquiring and processing verb argument structure : Distributional learning in a miniature language. *Cognitive Psychology 56,* 165–209.

Wright, J. (2000). Perrault's criticism of the Cartesian theory of the soul. In: Gaukroger, S., Schuster, J., & Sutton, J. (Eds.) *Descartes' natural philosophy.* pp. 680 - 696. London: Routledge.

Yang, C. (2004). Universal Grammar, statistics or both? *Trends in Cognitive Sciences, 8,* 451-456

Zimmer, K. (1968). Review of Chomsky 1966. *International Journal of Linguistics, 34,* 290-303.

Zimmerman, M. (1968). Is linguistic rationalism a rational linguistics? In: Hook, S. (Ed.) *Language and Philosophy.* (pp. 198 - 207). New York: New York University Press.

[CSM I, II; CSMK III] The Philosophical Writings of Descartes, v. I, II, transl. J. Cottingham, R. Stoothoff, & D. Murdoch, and v. III, transl. J. Cottingham, R. Stoothoff, D. Murdoch, & A. Kenny (Cambridge: Cambridge University Press, 1984, 1985, 1991). References are to volume and page number.