

GENERATIVE LINGUISTICS AND LANGUAGE ACQUISITION

1. INTRODUCTION

2. STANDARD THEORY AND LANGUAGE ACQUISITION

2.1 PRELIMINARY REMARKS

2.2 STANDARD THEORY

2.3 THE LAD WITHIN STANDARD THEORY

2.4 LEARNABILITY MODELS GROUNDED IN THE STANDARD THEORY MODEL

2.4.1 DEGREE-2 LEARNABILITY

2.4.2 DEGREE-1 LEARNABILITY

2.5 QUESTIONS ABOUT LANGUAGE ACQUISITION WHICH THE STANDARD THEORY CANNOT ANSWER

3. PRINCIPLES AND PARAMETERS THEORY

3.1 THE MODEL AND ITS IMPLICATIONS FOR THE STUDY OF ACQUISITION

3.2 SETTING VALUES FOR PARAMETERS

3.2.1 THE PROBLEM

3.2.2 POSSIBLE ANSWERS

3.2.2.1 PARAMETER FIXATION AND MARKEDNESS

3.2.2.2 PARAMETERS HAVE A PRESET VALUE

3.2.2.3 PARAMETER SETTING AS A TRIGGERING PROCESS

3.3 PRINCIPLES AND PARAMETERS AND LEARNABILITY

3.3.1 CUE-BASED ACQUISITION AND DEGREE-0 LEARNABILITY

3.3.2 THE STRUCTURAL TEMPLATES HYPOTHESIS

3.3.3 THE TRIGGERING LEARNING ALGORITHM

4. ACQUISITION FROM A MINIMALIST PERSPECTIVE

4.1 NEW FOCUS OF INQUIRY

4.2 THE MINIMALIST PROGRAM IN A NUTSHELL

4.2.1 GENERAL ORGANISATION

4.2.2 FROM THE LEXICON TO PF AND LF

4.2.2.1 REMARKS ON PROJECTIONS

4.2.2.2 MERGE AND ACQUISITION

4.2.2.3 MOVE AND ACQUISITION

SUMMARY

FURTHER READING

GENERATIVE LINGUISTICS AND LANGUAGE ACQUISITION

Who does not look back at where he came from will not reach where he is going. (Tagalog proverb)

KEY POINTS

In this chapter you will learn about:

- the way in which the generative approach to the study of language departs from the behaviourist approach
- the way in which generative models of linguistic analysis contributed to an understanding of language acquisition
- how language acquisition was dealt with by the Standard Theory model
- how language development was conceptualised within the Principles and Parameters model
- language acquisition and the Minimalist Program

1. Introduction

This chapter offers a brief presentation of generative models of linguistic analysis with a focus on the sense in which they have contributed to an understanding of language acquisition.

The rise of generative linguistics, associated with the name of Noam Chomsky, represented a radical shift from 'behavior or the products of behavior to states of the mind/brain that enter into behavior' (Chomsky 1986:3), a change of perspective from behaviourism, which dominated the social sciences in the 1950s, to mentalism, which understands 'talk about the mind to be talk about the brain at an abstract level at which [...] principles can be formulated that enter into successful and insightful explanation of linguistic (and other) phenomena that are provided by observation and experiment' (Chomsky 1987:50). Within such an approach, the Cartesian idea that language is a mirror of the mind is resurrected. The main empirical assumption about language is that there is a specific faculty of the mind/brain that can account for the acquisition and use of language.

Obviously, such a view represented a significant shift from the school of thought of the well-known psychologists of the time (such as John Watson or B. F. Skinner) who rejected the study of mind as unscientific. The behaviour of organisms was explained with laws of stimulus-response conditioning. The organism (animal or human) was seen as an empty black box while the stimulus (or the input) and the response (or the output) represented the only objectively measurable entities:

Generative linguistics represented a shift from behaviourism to mentalism.

Language = a mirror of the mind.

Behaviourism:
(knowledge of) language
= (knowledge of) a set of
habits,
dispositions
and abilities.



The mind was seen as a *tabula rasa* before learning (children are born, on such a view, only with the general capacity for analogies and overgeneralization) and learning was described as a combination of association, analysis, analogy, etc. Language was explained in the same vein since it was assumed to be just another form of behaviour. Thus, language can be learned just like any other skill (dancing, playing the piano, etc.). On analogy with other learning processes, one would then expect children to achieve a different level of language knowledge.

Skinner's 1957 book *Verbal Behavior* represented an attempt to explain language without taking into account any meanings, ideas or grammar rules, i.e. anything that might be defined as a mentalistic event. Skinner believed that verbal behaviour could be controlled by the same external processes which were used to control the behaviour of rats or pigeons since, he claimed, 'the methods can be extended to human behaviour without serious modifications' (Skinner 1957: 3). The methods relied on classic conditioning. Imagine a hungry pigeon which is in a box. When it pecks a button by chance, it will receive food. After pecking the button on several occasions, the pigeon will come to learn the connection between the button and food. It will receive positive reinforcement every time it pecks the button: food is provided. Learning language is only one more type of conditioned learning by association. The first sounds a child utters are shaped up by reinforcement (of behaviour by means of rewards) to grow into the full range of verbal sounds of adult language. A verbal response is weakened or strengthened, depending on the type of consequences it may have: negative or positive.

Skinner provides a few examples of how verbal responses are conditioned and reinforced. A *mand*, for example, is, according to him, the result of need stimulus, such as the need for water. The verbal response to such a stimulus may be the message 'Give me some water'. The person being given what he/she has asked for reinforces the message: he/she says 'Thank you', which reinforces the response of the listener, and so on. When an adult teaches a child how to speak correctly, positive reinforcement is given by saying: 'That's right' or 'Good' when appropriate linguistic behaviour has been emitted. What happens when the behaviour is not the expected one? Skinner suggests... 'generalised punishments' (!). How do children create new sentences? Sentences are defined as strings of words, organised in linear order. Within the behaviourist approach, language is thus acquired by habit-formation, via positive/negative reinforcement.

When acquiring language, defined as a set of habits, gradually built over the years, the child must solely rely on environment. The study of language acquisition is reduced to the study of observables, i.e. of input-output relations, without resorting to any study of the internal structure of the organism.

In 1959, Noam Chomsky, in his famous critical review of Skinner's book, argued that the stimulus-response model is completely untenable for language behaviour. Firstly, such a system cannot account for the production and comprehension of entirely new sequences of words. We can understand/utter sentences which we have never heard before. Chomsky's famous sentence 'Colorless green ideas sleep furiously' clearly proves that any sequence of words which has not been heard before can, however, be recognised as a grammatical sentence. A stimulus-response model cannot possibly

Behaviourism:
the study of language
acquisition is reduced to
the study of
input-output
relations.

¹ The "empty black box" was later abandoned even by behaviourists. Neo-behaviourism argues in favour of the idea that the stimulus-response connection is not sufficient to deal with the problem of situations; there must be some internal mechanism that allows the organism to choose new responses when facing certain situations. The idea of internal mediating mechanisms was introduced. These mechanisms are assumed to account for the fact that the same stimulus does not always produce the same responses.

explain the fact that every sentence which a person might understand or utter can be a novel combination of words or that children can acquire language rapidly, without any formal instruction, growing to correctly interpret constructions they have never heard before. Language cannot be described as a repertoire of responses nor can language acquisition be defined as the process of learning this repertoire.

Such an approach to language acquisition can neither account for the lack of negative evidence. It has already been shown that communication between children and parents does not seem to depend in any way on the ill- or well-formedness of children's utterances. Parents only rarely correct their children's (grammatical) errors. According to Skinner's hypothesis, the child learns how to speak correctly via positive and negative reinforcement. The hypothesis does not say anything about what may happen in the absence of negative reinforcement and, consequently, it cannot answer the question of how children manage to acquire language in the absence of negative input.

The behaviourist view does not make any assumptions about the learner's predisposition to learn. It cannot explain why only human beings can acquire speech, if knowledge of language can be achieved via conditioning.

It also allows for an unstructured hypothesis space and thus renders the task of learning extremely difficult (Saleemi 1992) and slow. There is evidence that children learn both grammar and vocabulary rapidly, sometimes after one single exposure to a particular word or a particular structure. Behaviourism also oversimplifies the problem². It has been shown in Chapter 1 that language cannot be acquired merely by resorting to analogy or associations, i.e. to domain-general learning mechanisms.

The generative approach seeks to explain the problems raised above on the assumption (already discussed in Chapter 1) that the brain is equipped with a faculty, which allows children to build an infinite number of new sentences out of a limited, deficient input.

The central problems of the study of language are, within such an approach, the following ones:

- (i) what is the system of knowledge called 'language'?
- (ii) how does the child acquire this system of knowledge on the basis of a deficient linguistic input?

The answers provided by generative linguistics to the issues in (i) -(ii) above are crucially different from the ones provided by behaviourism. Language is no longer interpreted as a system of habits, dispositions and abilities but as a computational system of rules and constraints, specific to humans. Such a view on language obviously led the path to a radically different interpretation of how knowledge of language is attained.

The empty black box of early behaviourism is replaced by the language acquisition device (LAD) of the language faculty which is far from being 'empty'. It contains the tools which help the child to construct a correct steady output on the basis of the PLD and which are responsible both for the great speed with which humans acquire language as well as for their creativity. The LAD is regarded as the device with which the child is equipped from birth, it is the initial state of language.

Hypotheses about the contents of the device itself have varied from one model to another. As we are going to see, within a Standard Theory approach, it contains

² To imagine that an adequate grammar could be selected from the infinitude of conceivable alternatives by some process of pure induction on an infinite corpus of utterances is to misjudge completely the magnitude of the problem. (Chomsky and Miller 1963:277)

Generativism: the child is equipped with a LAD, which is responsible for the speed with which humans acquire language as well as for their linguistic creativity.

substantive and formal universals, within a Government and Binding approach or within a Minimalist one, it is defined as containing a set of principles and a set of parameters.

As access to the LAD is indirect, in the sense that we can only hypothesise about it relying on the data offered by the analysis of the input and of the output, it is but natural that the details varied from one model of grammar to another. The changes reflect the fact that a better understanding of the acquisition process can only be achieved by revisions in the linguistic theory. At the same time, investigating language acquisition, 'one may hope to give some real substance to the traditional belief that the principles of grammar form an important, and very curious, part of the philosophy of the human mind.' (Chomsky 1965:59)

Generative linguistics seeks to meet explanatory adequacy.

One of the goals of generative linguistics has been, from the very beginning, explanatory adequacy. Choosing one model or the other also takes into account the ability of the model to explain the process of acquisition. The linguist will prefer that particular model, i.e. that theory of grammar, which can best account not only for what languages share and for what distinguishes one language from another, but also for how children manage to learn language so fast, without any conscious effort. Developmental facts can be extremely revealing for the study of the organisation of the language system.

Within such a view, knowledge of language is no longer interpreted as relying on analogy and it is regarded as a different, specific skill. The use of language is defined as rule governed behaviour.

2 Standard Theory and Language Acquisition

2.1 Preliminary remarks

The Standard Theory (ST) of transformational generative grammar was first formulated in Chomsky's (1965) *Aspects of the Theory of Syntax*. From the point of view of acquisition it represents the first systematic attempt to formulate an argument in favour of a rationalistic approach to the study of language³ which assumes that 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 organised way'. (Chomsky 1965: 48).

Equally important, it raises the problem of adequacy (or adequacy-in-principle) and that of feasibility. The former concerns the matching of the principles (to be applied to primary data) proposed by the theory with the various grammars which are actually discovered when the linguist examines real, natural languages. The question of feasibility concerns the potential of a certain theory to explain how grammars are produced⁴. Accounting for the learning algorithm becomes a condition imposed on any linguistic theory.

³ In this respect, Chomsky follows the line of Descartes, Leibnitz or Humboldt. His point of view differs sharply from that of philosophers like Quine and Wittgenstein, who adopt the empiricist position with regard to language, which they assume to be independent of innate mental faculties and thus learnable through drill or explanation.

⁴ Wexler and Culicover (1980), for example, define feasibility as easy learnability, i.e. 'learnability from fairly restricted primary data, in a sufficiently quick time, with limited use of memory' (p.18).

2.2 Standard Theory

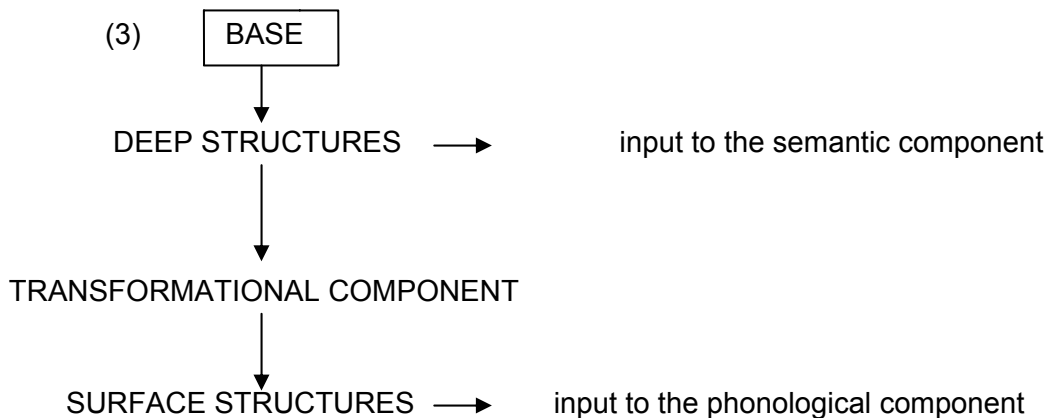
In order to understand what view of language acquisition the ST approach can offer and, more specifically, what the LAD is assumed to consist of, one should first examine the model of grammar adopted by ST. Remember that the LAD can only be hypothesised on the basis of the analysis of the input and of the output, i.e. its contents derive from the model of grammar assumed.

In what follows, the ST view on the organisation of grammar will be briefly presented⁵, with a focus on those aspects which are directly relevant for the present discussion.

A ST grammar is organised in various components. It consists of a base, which contains a set of context-free phrase structure rules and a lexicon. The lexicon is defined as a list of idiosyncratic properties of lexical and grammatical formatives; the entry of each item contains specification of its phonological and semantic properties, as well as information about its subcategorisation and selectional properties. The context-free phrase-structure rules rewrite⁶ single non-terminal symbols wherever they may appear in the derivation, i.e. they apply independently of context, and they are of the type shown in (2):

- (2) $S \rightarrow NP \text{ AUX } VP$
 $VP \rightarrow V \text{ NP}$
 $NP \rightarrow \text{Det } N$
 $\text{Det} \rightarrow \textit{the}$
 $PP \rightarrow P \text{ NP}$

The rules apply mechanically and in a stepwise fashion until no more rules can be applied. At this point, a structure has been created. The structures produced by the base, the deep structures or the underlying structures, are taken over by a transformational component (which, in its turn, consists of transformational rules) which maps them into surface structures:



The ST model: the LAD consists of substantives (grammatical categories, their projections, and grammatical features) and formal universals (formats constrained by UG).

The deep structures and the surface structures provide input to the semantic and the phonological components respectively. Notice that, within the ST model, only deep structure is subject to semantic interpretation.

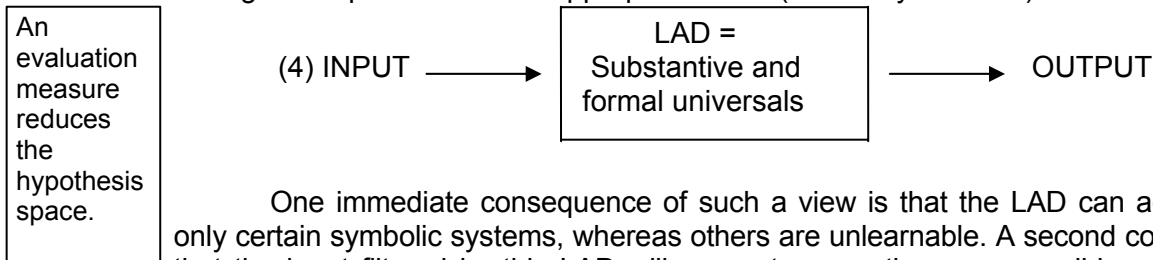
⁵ For a more detailed presentation of the ST model the reader is referred to Ruwet (1967), Chapter 2 in Jerban (1982) or Chapter 2 in Cornilescu (1995).

⁶ That is why they are also called rewriting rules.

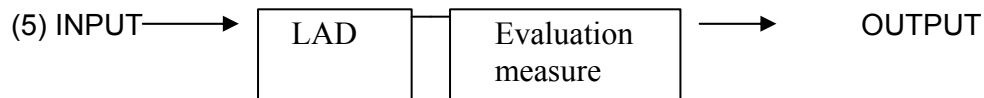
Each level of representation is derived from another and derivation is mediated by rules. These rules are defined on sets of substantive universals, i.e. grammatical categories (N, V, P, A) and their phrasal projections (VP, VP, PP, AP) and grammatical features (+V, +N) in the syntactic component, semantic primitives in the semantic component (such as [+/-abstract], [+/-human], etc. and phonological features in the phonological component⁷. Rules are construction-particular. There are rules such as Dative Transformation, Relative Clause formation, Passive formation, Reflexivization, a.s.o. They are composed of elementary operations and have different formats, which are constrained by formal universals⁸.

2.3 The LAD within ST

The LAD, a mediator of the input-output relation, is defined as consisting of these substantives and formal universals which 'provide a schema that is applied to data and that determines in a highly restricted way the general form [...] of the grammar that may emerge from presentation of appropriate data' (Chomsky 1965:53):



One immediate consequence of such a view is that the LAD can acquire or use only certain symbolic systems, whereas others are unlearnable. A second consequence is that the input filtered by this LAD will generate more than one possible grammar. How does the child choose one single grammar out of this set? The acquisition process is assumed to involve an *evaluation measure* which imposes a ranking on the members of the set of possible grammars, thus reducing the hypothesis space and allowing the child to choose that grammar which is the most compatible with the data offered by the input and which has the status of 'predicted descriptively adequate grammar':



Most subsequent studies of language acquisition (Berwick 1985, Atkinson 1992, Saleemi 1992, O'Grady 1997) criticised the ST model for offering an instantaneous view of the acquisition process, i.e. all the data seem to be accessible to the child at once. However, several remarks are in order here. Firstly, one should not ignore the fact that such an idealisation is legitimate from the point of view of linguistic theory and that it was necessary when the focus was on showing that language acquisition mirrors mainly 'the general character of one's capacity to acquire knowledge - in the traditional sense, one's innate ideas and innate principles' (Chomsky 1965:59) and not 'so much the course of

⁷ Substantive universals could be defined as 'primitive elements which a grammar establishes in order to analyse linguistic data' (Crystal 1985:295) and which consist of 'any feature or category, phonological, syntactic or semantic, which is part of the vocabulary necessary and sufficient for the description of the world's languages' (Smith and Wilson 1979: 288).

⁸ They specify 'the form of rules in a grammar' (Smith and Wilson 1979:253.) representing 'the necessary conditions which have to be imposed on the construction of grammars in order for them to be able to operate'. (Crystal 1985 : 321).

one's experience'. Chomsky (1965: 202) himself stresses the fact that he is idealising and points out the developmental dimension which is involved in the process:

Obviously, to construct an actual theory of language learning, it would be necessary to face several other very serious questions involving, for example, the gradual development of an appropriate hypothesis [...] and the continual accretion of linguistic skill and knowledge [...]. What I am describing is an idealization in which only the moment of acquisition of the correct grammar is considered. [...] it might very well be true that a series of successively more detailed and highly structured schemata (corresponding to maturational stages but perhaps in part themselves determined in form by earlier steps of language acquisition) are applied to the data at successive stages of language acquisition.

This quotation alone proves that the idea of language growth is present in the model as early as 1965, and that the criticism that the schema for language acquisition offered by the ST model is characterised by a 'total lack of developmental dimension' (Atkinson 1992:43) is not grounded, if not misleading. It is one thing to say that the model which it offered was idealised, and hence language acquisition was presented as instantaneous, but it is quite a different thing to say that the model lacked a developmental dimension. As we are going to see, the learnability models grounded in the ST framework assumed an incremental view of the process of language acquisition.

The model of language acquisition which ST could offer was faulty to the extent to which the model of grammar was faulty, i.e. it was not the abstract way in which language acquisition was conceived that was at stake, but the type of hypothesis one could build about the LAD on the model of grammar which the linguistic theory could offer at the time. The general assumptions which lie behind the history of language acquisition within generative linguistics have remained the same. What has been changing, in an attempt at gaining a better understanding of the language faculty, has been the descriptive, not the explanatory, part of the grammatical theory. That language acquisition is a gradual process, that it represents a development of language skills which relies on some species-specific innate predispositions is an idea which has always been present in generative studies. It is the model of grammar which has been revised, very often with the goal of leading towards a more appropriate learnability theory.

It should also be pointed out that acquisition of language, though associated with a distinct specific faculty of the human mind, is not seen as entirely separate from other components of the abstract mind:

Notice that we do not, of course, imply that the functions of language acquisition are carried out by entirely separate components of the abstract mind or the physical brain, just as when one studies analyzing mechanisms in perception [...], it is not implied that these are distinct and separate components of the full perceptual system. In fact, it is an important problem for psychology to determine to what extent other aspects of cognition share properties of language acquisition and language use, and to attempt, in this way, to develop a richer and more comprehensive theory of mind. (Chomsky 1965:207)

2.4 Learnability models grounded in the ST model

2.4.1 Degree-2 Learnability

One of the first attempts at providing a theory which unifies linguistic representation and learning is associated with the names of Ken Wexler and Peter Culicover. In their 1980 book, *Formal Principles of Language Acquisition*, they tried to show that transformational grammar is 'naturally' and 'easily' learnable⁹, i.e. they define 'feasibility' as 'easy learnability', 'that is, learnability from fairly restricted primary data, in a sufficiently quick time, with limited use of memory' (Wexler and Culicover 1980:18), and developed a theory of language learning which incorporates a theory of generative grammar and a learning procedure.

They start from two main assumptions:

- (i) the primary data provided by the input which the child receives consist each of a base phrase-marker (b) paired with a surface string (s): (b,s). The base phrase marker (or the deep structure) is conceived as close to the semantic representation of the utterance. The child can understand a sentence even when he/she cannot analyse it because he/she relies on the situation in which the sentence is uttered as well as on the spoken words. Thus, the child is assumed to have the capacity to relate this interpretation to syntactic structures, i.e. semantic information facilitates access to deep structures;
- (ii) the child's task in the process of language acquisition is to construct a transformational component.

The learning mechanism which they propose is quite simple and gradual: it creates transformations on the basis of PLD (i.e. pairs of base phrase markers and surface strings). The child's task is that of learning the transformational rules. Whenever the input datum allows an error to be detected (i.e. whenever the transformation hypothesised by the learner is different from the transformation in the target grammar), that transformation is deleted and the child will hypothesise a new transformation. It is important to point out that the mechanism can 'see' only the latest datum, it does not go back to earlier data; the rejection and the selection of transformations does not target the transformational component wholesale. Hypotheses are changed gradually, as new data are provided by the input.

The main idea is that the child can select a correct transformational component on the basis of a relatively simple input. Even though eventually the learner will master a grammar which contains complex sentences, in the process of language acquisition the most complex phrase marker which the child must consider will contain no more than two embeddings. Hence the name of the theory: degree-2 learnability. In the process of acquisition, the child will make some errors which, for the learning to take place, must be detectable errors.

Within the framework adopted, all the transformations are obligatory. Thus, the child will be able to see that his/her hypothesised transformation is incorrect every time the input provides a paired base structure phrase marker and surface string if the surface string is different from the one his/her transformational component would have

Degree-2 Learnability: on the basis of a simple input, the child constructs the transformational component, considering phrases which need contain no more than two embeddings.

⁹ 'Learnable' should not be understood to mean learnable by some general learning mechanism which can account for the learning of any kind of cognitive capacity. As Wexler (1982:286) points out: 'an ability (e.g. the ability to speak a natural language) is learnable if there is an (empirically true) way in which the ability can develop. In this sense, if an ability is innate it is learnable [...]. If the ability develops via a learning mechanism based on other innate capacities, the ability is also learnable [...]. In short, any human capacity is learnable.'

generated from the same base structure. Suppose the child's transformational component has generated the surface string *s1* from the base structure *b*. Then the child hears a primary datum (*b*, *s2*) and realises that the transformation he/she applied to *b* is incorrect. On the basis of the latest datum, a new transformation will be hypothesised. But some errors may not be detectable until a degree-2 phrase marker. The hypothesis is that transformational grammar 'can be learned from data of degree less than or equal to 2' (Wexler and Culicover 1980:117), i.e. errors are detectable on a base phrase marker of degree-2 or less. That would be in line with the assumption that the input which a child receives is relatively poor; however, grammar can be learned from an input of reasonably small complexity: 'Nothing more complicated than sentences that contain sentences that contain sentences is needed for convergence' (Wexler 1982:299).

The transformation theory provided by ST was relatively unconstrained. Hence, one of the major problems for degree-2 error detectability was the possibility of a low cycle error to be undetected until on a much higher cycle. That required phrase markers of a high degree of embedding as part of the PLD, an undesirable result. Given this shortcoming of ST, Wexler and Culicover had to add a number of constraints, conditions and principles to the standard definition of transformations¹⁰.

Their work shows in what way developing a learnability theory is linked to the development of linguistic theory itself. As Baker (1982) puts it: 'only by making major revisions in the linguistic theory itself were they able to achieve satisfactory learnability results' (p.420).

However, the model they propose is not without problems. Most of the constraints they propose are 'specific', providing a solution to one kind of situation. Assuming that all the transformations are obligatory, their model cannot give an account of error detectability in the case of those errors linked to optional rules (although they assume many of the rules which were formulated as optional rules).

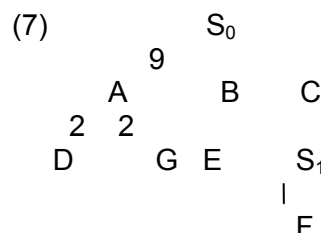
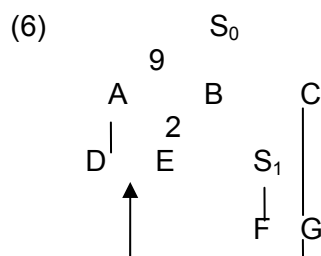
Also, recall that one of the background assumptions they made was that the child has the ability to construct base structures on the basis of semantics. Such an assumption is, as Atkinson (1992) points out, both too strong and too weak at the same time: the input is defined as consisting of surface strings which do not provide any structural information.

2.4.2 Degree-1 Learnability

Morgan (1986) proposes an alternative learning theory which modifies the model of Wexler and Culicover by introducing the so-called Bracketed Input Hypothesis. According to this hypothesis, input is bracketed: it contains sequences of base structure – surface bracketing pairs, with certain clues about the way in which words are grouped into syntactic constituents. Suppose that the target grammar contains a transformation which operates on the base structure in (6) yielding (7):

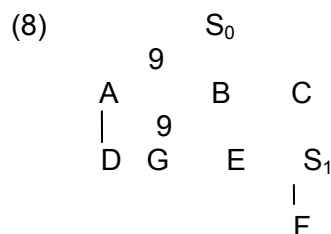
¹⁰ Among which the so-called Freezing Principle ('if a transformation changes the structure of a node so that it is no longer a base structure, nothing under that node may be analysed (and thus changed) by a further transformation' -Wexler and Culicover 1980:119), the Raising Principle (if a node is raised, no further transformations can apply to any of the nodes dominated by the raised one) or the Binary Principle ('a transformation may apply to the level at which it is operating plus the next lower level, but no further down' Wexler and Culicover 1980:109). These principles are still at work in more recent generative models, under different names.

Degree-1 Learnability: on the basis of an input which contains sequences of base structure – surface bracketing pairs, the child builds the transformational component considering phrases which need not contain more than one embedding.



As shown in (6) – (7) G raised and adjoined as a right sister to D.

Now suppose that the child has hypothesised a transformation that raises G but adjoins it as a left sister to E, as in (8):



The surface string in (7) and (8) will be the same: DGEF. The error will not be detectable if the input provides only strings (as in Wexler and Culicover's model) not structures. If the Bracketing Input Hypothesis is assumed, (7) and (8) will have different representations, as shown in (9) and (10) respectively:

(9) [D – G] [E] [F]

(10) [D] [G – E] [F]

The child will be able to detect the error on a phrase marker which contains no more than one embedding, i.e. on a degree-1 datum.

One can notice that Morgan enriches the input by adding 'bracketed' structural information to the simple input of Wexler and Culicover (1980), but then simplifies it by reducing the degree of complexity of phrase markers necessary for error detection and hence for learning.

Examining the two learnability theories from a contemporary stance, when constraints are generally limited to elements within the same clause, one might wonder why children need to hear more than a simple, unembedded clause in order to be able to construct a correct grammar. In 3.3.2 an alternative to the Degree-2 and Degree-1 theories will be presented: Degree-0, put forward by Lightfoot (1989, 1991), according to which everything can be learned from unembedded domains. Children do not seem to need access to complex structures in order to reach the correct generalisations.

2.5 Questions about language acquisition which ST cannot answer

The ST model was not restrictive enough in terms of the rule system.

In spite of the fact that the ST model (in particular the program set forth in *Aspects of the Theory of Syntax*) emphasised the central role of language acquisition in linguistic inquiry, it could not offer a satisfactory account of the process of acquisition. What should be stressed once again is that the model needed changing, among other things, because it could not deal in a satisfactory way with learnability phenomena.

The grammatical theory proposed by ST was not restrictive enough in terms of the range of grammatical rules which were allowed. Hence, the need to add constraints on rules¹¹ in order to provide an appropriate learnability theory (see Wexler and Culicover 1980, for example). The model offered too many types of transformation, whose ordering was (often) not available from the relatively impoverished and deficient input. This made it difficult to explain how the child could choose one particular grammar rather than another.¹² For example, the Passive Transformation Rule, formulated as in (11), raises several questions. How could a child detect, on the basis of the input, i.e. of surface strings (as in Wexler and Culicover's model) or surface structured information (as in Morgan's learnability model) the number of movements the transformation implied or the order in which the movements must apply: move first the two NPs (from position 5 to position 2, and from 2 to 7), insert *be* and *-en* and then insert *by* in front of the NP which has moved to the right?

(11)	X	-	NP	-	AUX	-	V	-	NP	-	Y	-	by	-	Z	
	1		2		3		4		5		6		7		8	→
	1		5		3+be		4+en		Ø		6		by+2		8	

The ST model also resorted to an evaluation measure (whose exact nature was left unspecified) which was designed to impose a ranking of the members of the set of possible grammars, i.e. to reduce the hypothesis space. However, the concept of evaluation measure was not only a sort of idealisation needed to account for the fact that, in the end, the child chooses one particular grammar, i.e. the one which is consistent with the input he/she has received, but it also heavily relied on the assumption that the child, given the LPD and UG, has to choose out of a set of candidate grammars. Within the Principles and Parameters (PP) model, where UG is constrained in such a way that the set of possible grammars is reduced, at all stages, the evaluation metric has been abandoned.

Questions about the role of the evaluation metric have also come from the domain of language change: *If the data and the associated structural description to which the child is exposed correspond fairly closely to the grammatical capacity of some older individual, one would expect the child's evaluation metric to select the same grammar as that of older individual's* (Lightfoot 1997). However, languages are not stable. If children acquire language without error (in the end), how does language change occur? What would the role of the evaluation metric be in this respect?

It is also worth mentioning at this point that the ST model does not say anything about conceptual learning, the arguments focussing exclusively on the learning of syntax. That triggered a lot of misunderstanding as well as criticism which targeted an area about which the model did not actually say anything.

¹¹ See Braine (1971) or Baker (1979) for comments on the disadvantage and inappropriateness of resorting to restrictions into the statement of rules.

¹² 'In essence, the problem is that there are too many rule systems. Therefore, it is hard to explain how children unneringly select one such system rather than another' (Chomsky 1987:15). 'Virtually any imaginable rule can be described in transformational terms. Therefore a critical problem in making transformational grammar a substantive theory with explanatory force is to restrict the category of admissible phrase markers, admissible transformations, and admissible derivations.' (Chomsky 1972:124).

3. Principles and Parameters Theory

3.1 *The model and its implications for the study of acquisition*

Much contemporary research in the acquisition of syntax has been carried within the framework provided by the theory which was first known as the Theory of Government and Binding (GB) (Chomsky 1981). As one of the main shortcomings of the ST approach resided in its failing to offer a theory of UG sufficiently constrained in the options it permitted and sufficiently devoid of stipulations that did not follow from general principles, which implied also failing to account for the fact that grammars can develop in spite of the limited input, the goal of the new approach was to provide a theory of grammar which should be able to meet both the descriptive adequacy and the explanatory adequacy conditions. Its descriptive statements fall into two categories: language-invariant and language-particular. The language-invariant statements are fundamental principles which can restrict the class of possible grammars and also constrain their form. From the perspective of acquisition, that meant reducing the searching space significantly. Also, with a general system of principles replacing the construction-particular rules which contained arbitrarily ordered operations, the model can better account for the acquisition process. The language-invariant statements also contain parameters, whose value is fixed by experience. The particular values which parameters have represent the language-particular statements and they account for cross-linguistic variation. The model assumes that UG provides a finite array of parameters with a finite number of available values, possibly only two.

As government and binding represent just two of the concepts assumed in the theory, and as the theory actually focuses on the interplay between the principles and the parameters provided by UG, the name of Principles and Parameters Theory (P&P), which captures the core of the model in a more accurate way, replaced that of GB.

This is not the place for a detailed presentation of the model¹³. P&P developed over more than a decade, with far-reaching ideas being pursued in many areas of linguistic theory, language acquisition included. I will only focus on the core ideas, trying to point out in what way the new model overcomes the shortcomings of ST. Needless to say, the emphasis will be on the consequences for the study of the acquisition process.

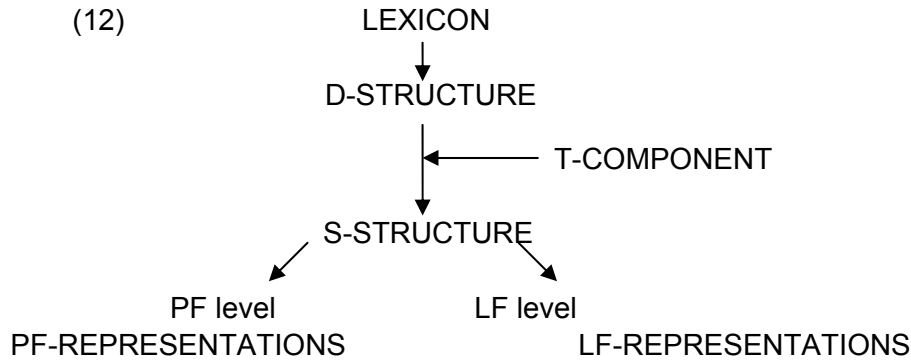
The overall organisation of grammar¹⁴ within a P&P approach is the one given in (12) below:

Within a P&P model, UG is defined as consisting of general principles and a set of parameters whose value is fixed by experience.

The searching space is reduced: the principles restrict the set of possible grammars and constrain their form.

¹³ For a detailed presentation of the theory, the reader will benefit from consultation of Lasnik and Uriagereka (1988), Haegeman (1991/1994), Chapter 1 in Chomsky (1995) or Cornilescu (1995).

¹⁴ One thing which should be made clear from the very beginning is that organisation of grammar within a P&P approach actually means UG. The concepts of 'grammar' and UG refer to two different systems: UG is an element of biological endowment and it characterises the child's pre-linguistic initial state. 'Grammar' is a certain system of knowledge, acquired on the basis of experience via UG, i.e. it represents experience filtered /mediated by UG. A particular grammar results from the process of fixing values to parameters.



The base consists of a lexicon (which is, just as its ST predecessor, a repository of idiosyncratic information about particular items) and a categorial component.

The categorial component within a ST approach contained context-free phrase-structure rules. But they could not capture the generalisation that syntactic categories have similar structure. Moreover, from the perspective of acquisition, postulating these rewriting rules in the base could not theoretically prevent the child from having to analyse strings of the type in (13) in his/her searching space, since nothing in UG excluded such rules:

- (13) a. $VP \rightarrow V - V - V$
 b. $PP \rightarrow P - P - P$

Of course, we can always assume that such rules do exist in the hypothesis space and that it is part of the child's task in the acquisition process to discard them and choose the 'right' ones in the end. However, that would mean placing a heavy burden on the young learner; the acquisition process would slow down in a significant way. The P&P categorial component contains rules that meet the requirements of X-bar theory according to which all phrases are headed by a lexical or functional head, they are endocentric. The head represents a zero projection (X^0) and the general format for phrase structure is the one in (14) below:

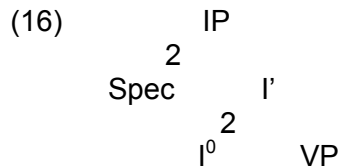
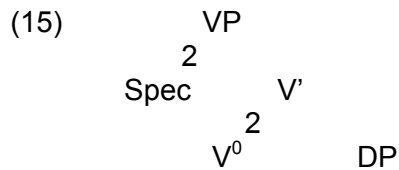
- (14) $X'' \rightarrow \text{Spec} - X'$
 $X' \rightarrow X^0 - (YP)$

The format in (14) captures what is common for the different types of phrases (VP, NP, AP or IP¹⁵, DP, etc.), providing a universal schema for phrase structure. It is part of UG. It is essential to point out that lexical (or open class elements) (15) and functional categories (belonging to closed classes)¹⁶ (16) project according to the same format:

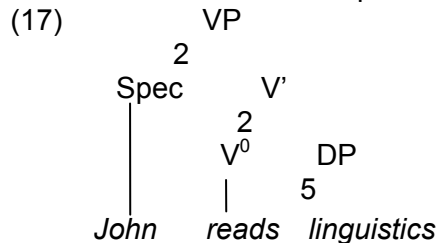
Lexical and functional categories project according to X-bar schema, part of UG.

¹⁵ IP = Inflectional Phrase; Inflection is an umbrella term for functional heads such as Tense and Agreement (Pollock 1989) or, within more recent studies, for all the functional projections of VP (Tense, Agreement Object, Agreement Subject, Aspect and Mood) and it represents the core of the sentence, being responsible for the relation between the subject and the predicate.

¹⁶ For more on the properties of functional categories see Abney (1987), Grimshaw (1991), Avram (1994) or Cornilescu (1995).



Both lexical (verbs, nouns or prepositions, for example) and functional categories (such as Tense, Agreement or Aspect) can be the *head* of a phrasal category (in 15 above V^0 , a lexical category, is the head of VP and in 16 Inflection – I^0 , a functional category, is the head of IP), taking a complement and a specifier. For example, V, the head of the phrasal projection VP, takes a determiner phrase (DP) as its complement, whereas the Specifier position is assumed to be the base position of the subject DP:



Now, the child will no longer have to discard rules like the ones in (13) for the mere reason that such rules are not possible in the given (X-bar) system provided by UG. There is one single available format for both lexical and functional material.

The lexicon contains relevant information about items, be they substantive or functional. Each entry will contain only idiosyncratic information, i.e. information which is not already provided by UG or which cannot straightforwardly follow from UG. Thus, a typical entry consists of information about the phonological matrix, about categorial status (whether the entry is + N or + V, for example) as well as phi-features (number, gender, person, and Case). Let us see how this works in order to understand its implications for the acquisition process. Consider a verb like 'read'. In the lexicon, 'read' comes together with information about its phonological form, a specification about its syntactic category (it is +V, -N) and with information about its notional (or thematic) structure, i.e. information that it takes an Agent, the reader, and a Patient, the object which is being read:

(18) read V: Agent, Patient

The association between these notional arguments and argument positions in syntax seems to be largely predictable (thematic properties 'specify' the argument structure) and will not be stated in the lexicon. The representation in (18) is then projected to D-Structure:

(19) NP1 read NP2

The relation between the lexicon and D-structure is a principled one: the information in the lexicon is projected to all the levels: D-Structure, S-Structure and LF. The subcategorisation properties of lexical items must be observed at each syntactic level. This is the so-called Projection Principle, which prevents, among other things,

The lexicon contains lexical and functional items. Each entry contains idiosyncratic information about the phonological matrix, categorial status and phi-features of the particular item.

The relation between the lexicon and D-Structure is governed by the Projection Principle.

strings of the type in (13). Notice that the representations at the level of D-Structure (an internal interface level where lexical properties are expressed in a form accessible to the computational system) are determined by the Projection Principle and not independently generated by phrase-structure rules¹⁷, as in ST, which become redundant within this model. Both subcategorisation and selectional restrictions follow now from theta-role specification and from thematic properties. The semantic features of an argument are required to be compatible with those of the theta-role associated with it. The D-Structure representations are the projection of the information stored in the lexicon, i.e. the projection of lexical properties. The thematic positions are saturated at this level. The Projection Principle and the X-bar format are given by UG. The child has to learn the idiosyncratic properties of lexical items. Chomsky (1981) did not specifically address the question of how exactly children learn these properties. But the suggestion is that knowledge of subcategorisation properties determines knowledge of the syntactic properties of that specific item. For example, an English speaking child must discover the subcategorisation features of *read*, one aspect of learning its meaning. Given this knowledge, the properties of the syntactic structures in which *read* appears are determined by the Projection Principle, which means that they do not have to be learned independently. Similarly, knowledge of the lexical properties of an item will lead to knowledge of its LF representation. The implications of this view, according to which subcategorisation frames are derivative from notional properties, are far from trivial and they have far-reaching consequences especially for the study of lexical development.

The most radical consequence would be to assume that once the child knows the semantic properties of a word, he/she would also be able to cope with the syntactic frame of that particular word. Another non-trivial consequence of this line of reasoning is that semantic primitives, of the type Agent, Patient, etc. can be applied by the learner in his/her pre-linguistic analysis of the input, whereas notions such as NP, VP are derived. The former would meet, on such a view, the condition of epistemological priority, representing the basis for linguistic development. From the perspective of acquisition, this view raises at least one important question: how does the child know that an element in a sentence is an Agent or a Patient before having any knowledge of syntax? For acquisition to be possible, it may be the case that both information about the thematic structure and about the argument structure are available in the input at the same time (Chomsky 1995). However, notions such as 'Agent' belong to the conceptual compartment whereas argument structure belongs to the computational compartment.

D-Structure is no longer the feeder of information to the semantic component, as in the ST model. Surface structure, associated with the external interface LF, becomes crucial for semantic interpretation: the semantics can be read off surface strings. This creates a tension between the assumption that thematic information determines argument structure and the view that semantic information is accessible via surface structures. In the domain of lexical development, the consequence of this tension is reflected in the by now classical debate over 'what comes first: syntax or semantics?'

¹⁷ For a discussion on whether phrase structure rules are/are not necessary within GB see Lasnik and Uriagereka (1988), pp. 2–5.

D-Structure representations are mapped to S-Structure via a basic operation Move α .

Move α always has the same format and its application is constrained by general principles. The child's task is thus easier.

The D-Structure representations are mapped to S-Structure by rules of the transformational (T) component and then they 'branch' to PF and LF. Recall that within a ST approach there were too many construction-specific rules, such as Passive, Question Formation, Dative-Shift, a.s.o. The acquisition task was rendered extremely difficult not only because there were too many transformation rules which the child was supposed to acquire but also because these rules were decomposed into more elementary operations which were said to apply in an order which was often an arbitrary one and which was not given by the input. Within a P&P approach, the transformational component consists of one single basic operation Move α , where α is arbitrary. Deletion and insertion are also operations of this component.

Move α is a general rule, which says that anything can be moved, leaving behind a trace, co-indexed with the moved element. The application of Move α is constrained by conditions imposed by general principles, which are applied to construct or license derivations in order to prevent the rule from producing a wide range of S-structures.¹⁸ The child is no longer supposed to learn a lot of transformation rules, each with its own more elementary operations. Move α always has the same format. The trace which it leaves behind represents an empty category, i.e. one that lacks a phonological matrix, but which is relevant for the operations of the computational system¹⁹.

The sub-components of the rule system, i.e. the lexicon, the syntax, the PF-component and the LF-component, interact with the principles of UG which are operative in all or many of the interacting modules of grammar which license different types of information at different stages of the derivational process: (i) government theory; (ii) binding theory; (iii) theta-theory; (iv) bounding theory; (v) Case theory; (vi) Control theory. The properties of languages are derived from the interaction of the sub-systems of the rule system and the sub-systems of principles.

The LAD is assumed to provide a set of universal principles (i.e. a fixed network) and a set of parameters²⁰ (a switch box) whose value will be fixed on the basis of input:



The principles provided by UG reduce the hypothesis search in a significant way. Acquisition of the computational system reduces to the process of fixing the parameters of the initial state. The task of setting the correct value for parameters is made easier, since certain forms or configurations are simply impossible given the system of principles of UG. The primary task of the child is to set values to parameters.

Within such an approach, language acquisition is defined as 'the acquisition of lexical items, fixing of parameters, and perhaps maturation of principles' (Chomsky 1995:28). The question is what the exact nature of 'fixing of parameters' is, what in the input can license parameter setting. What is the role which maturation plays in the process?

Language acquisition = the acquisition of lexical items and fixing the value of the parameters of the initial state.

¹⁸ Just like in the case of d-structures (which should not be mistaken for the deep structures within the ST approach), S-Structure is not the P&P equivalent of the ST surface structure. Recall that in ST the input to the semantic component was deep structure. The surface structures were not linked to semantic interpretation. In the schema given in (6) one can easily notice that within a P&P approach S-Structure provides input to LF.

¹⁹ For an example of how such a trace may work and of how children deal with traces very early, see the discussion on *wanna* constructions in Chapter 1.

²⁰ The parameters may have both settings (+/-) available (Valian 1990, Fodor 1998 a.o.) or they may be available with a default setting (Hyams 1986).

Parameters are mainly associated with functional categories, such as Determiners or Tense, which represent the locus of language variation. If UG provides the general principles, one may assume that the child has to learn only what is language-specific, i.e. the values associated with functional categories. Recall that functional categories project just like lexical categories. An element is assumed to come from the lexicon in its 'bare' form and then, via head-to-head movement, moves to the functional categories in its extended projection in order to get its inflection (see, for example, 49). This view allowed building-structure accounts of syntactic development (Lebeaux 1988, Radford 1990, among others), according to which the child's phrase marker 'grows' as the child sets the right values for the parameters associated with functional categories.

3.2 Setting parameter values

3.2.1 The Problem

Reducing the task of acquisition to the search for the correct parameter values also meant reducing the search space. However, various questions were raised with respect to the nature of the process of parameter setting:

- (i) in what order does the young learner consider the values of a given parameter?
- (ii) does UG provide an initial set of parameter values?
- (iii) does the child fix the correct value from the very beginning or does he/she reset the parameter on the basis of positive data?
- (iv) is the process of parameter fixation best defined as hypothesis formation and testing or as triggering?
- (v) what element(s) in the input can count as triggers for parameter setting ?

Let us address these questions and discuss some of the answers that the literature can offer at the moment.

3.2.2 Possible answers

3.2.2.1 Markedness and parameter fixation

The answer to question (i) may rely on the notion of markedness, which has been analysed as playing the part of an evaluation metric in the process of parameter fixation. One possible answer along this line is the one provided by the so-called Subset Principle, versions of which have been proposed by Baker (1979), Berwick (1985), Manzini and Wexler (1987), Wexler and Manzini (1987), Wexler (1993). In what follows I will only refer to this principle in the form in which it was formulated by Manzini and Wexler.

Their argument has been developed in relation to binding domains and it relies mainly on the analysis of the interpretation of reflexive pronouns. Languages differ with respect to the constraints imposed on the co-indexation of reflexive pronouns with their antecedents. In languages like English, for example, a reflexive pronoun can only have an antecedent in the same clause. 'Himself' can be co-indexed with 'John', in (21a), since they are both in the same clause. But in (21b), 'himself' can no longer be co-indexed with 'John', since 'John' is in a higher clause:

- (21) a. *John_i admires himself_i.*

The process of parameter fixation may be linked to markedness

b. **John_i believes [that Bill admires himself_i]*.

In other languages, such as Japanese or Korean, reflexive pronouns can be co-indexed with an antecedent that is either in the same clause or in the higher clause, provided the reflexive and its antecedent are in the same sentence. Thus, Principle A of Binding Theory (formulated in 22) can be parameterised, with values (a) and (b) in (23) as possible options:

- (22) Principle A:
An anaphor must be bound in its governing category.
- (23) The governing category may be:
 - a. the smallest clause which contains the anaphor (the reflexive pronoun)
 - b. the sentence containing the anaphor (the reflexive pronoun).

One main assumption is that each parameter has to observe the so-called Subset Condition:

- (24) For every given parameter and every two given values of it, the languages defined by the values of the parameter are one a subset of the other. (Wexler and Manzini 1987:60)

This set inclusion is regarded as determining the ordering of the two values and hence the ordering of parameter fixation: the child will begin with that value which allows the smallest range of structures and, on the basis of experience, he/she will then opt for that value which allows for a wider range of constructions. This is the so-called Subset Principle which could be formulated in simple terms as in (25):

- (25) If every structure that is grammatical under A is also grammatical under B, then choose option A if it is consistent with the input. (Wexler 1993:217)

Let us go back to Principle A and see how the principle works. The child acquiring English will come across sentences like the ones in (26), where the reflexive pronoun can be co-indexed with an NP which is once in the smallest clause (26a) and once in a complex sentence (26b):

- (26) a. *John admires himself.*
b. *I think John admires himself.*

Recall that the Subset Principle says that the LAD must choose the most restrictive value of the parameter, provided it is consistent with the positive data which the child has received. Hence, the child will choose option (a), according to which the reflexive pronoun is co-indexed with its antecedent in the smallest clause. This choice allows for a more restrictive range of constructions. In this way, choosing the value which permits a wider range of constructions is disallowed. From the point of view of learnability, there will be no need for correction. If the child chose the value consistent with the wider range of constructions, there would be no way to ever retreat to the correct value, since the positive data provide sentences which are consistent with both options.

It is important to point out that the Subset Principle is viewed as an independent principle, part of a learning module which interacts with UG. The non-trivial implication is that the child resorts to an explicit learning procedure and that acquisition theory needs more than linguistic theory (vs. the view advanced by Hyams 1986). The second implication of this view of parameter setting is that markedness plays an important part in

The ordering of parameter fixation must observe the Subset Principle: choose the most restrictive value of the parameter which is consistent with the positive data.

The Subset Principle is an independent principle, part of a learning module which interacts with UG.

the ordering of parameter values: parameter fixation starts with the unmarked value and then, on the basis of positive data, moves to the marked one.²¹

A critical look at the set-theoretical approach of Wexler and Manzini will reveal that it is not without problems. Saleemi (1992) qualifies it as 'far too deterministic to be psychologically convincing and too conservative to be applicable to a wide range of parameters' (p.76).

Gair (1988) points out that assuming that there are marked and unmarked parameters which determine core grammar can have far reaching implications for language acquisition and language change: (i) languages can be ranged along a scale of markedness, according to the number of marked/unmarked parameters which determine their core grammar; (ii) one would expect marked languages to move towards unmarkedness and marked languages to be more stable (fact which has not been borne out by empirical data); (iii) some languages (the 'unmarked' ones) can be acquired faster than others, since the unmarked values are set first.

Recall that the Subset Principle relies on the Subset Condition. But it is not at all clear that all parameters observe the Subset Condition. Take for example the head-complement order parameter. The two values of this parameter are (a) and (b) below:

- (27) a. the head precedes the complement (i.e. head-initial)
- b. the complement precedes the head (i.e. head-final)

Both (a) and (b) are equally restrictive, neither of them is a subset of the other. In this case, the Subset Principle cannot apply. Obviously, as O'Grady (1997:284-285) points out, 'this need not create a learnability problem (since an incorrect setting would quickly run up against counterexamples in experience), but it does suggest that the parametric options permitted by UG are not all of the same type'.

Another parameter which questions the Subset Principle is the so-called null subject or *pro*-drop parameter which can have two values:

- (28) a. null subjects are allowed.
- b. null subjects are not allowed.

Languages like Romanian, Italian or Spanish will select value (a) which allows for a wider range of constructions (i.e. for both structures with null subjects and structures with overt subjects) whereas languages like English select value (b). With value b being more restrictive, according to the Subset Principle the child acquiring English should never use null subject sentences, since he/she should first opt for the b value of this parameter which is also the correct value for the target language. But empirical data from child English show that children choose value (a) first.

²¹ The idea that markedness is a sort of evaluation measure is explicit. See for example the definition of the Subset Principle in Wexler (1993:219):

Subset Principle (Relativized Statement): Suppose that principle X allows for 2 values, i and j, of a parameter P. Suppose furthermore that for all derivations D, if D(j) violates X, then D(i) violates X. Then value i of parameter P is unmarked with respect to value j.

3.2.2.2 Parameters have a preset value

Parameters have one value set prior to experience with a given language. The child has to reset this value on the basis of positive data.

Another possible answer with respect to the ordering of parameter fixation is the one provided by Hyams (1986). Markedness is no longer considered an evaluation device whose task is to determine which parameter value is the non-marked one and hence the priorities in the process of parameter setting. Hyams's theory is based on the analysis of null subjects in child English. The main pillar of her evidence is that children acquiring English use null subjects where adults would require an overt one. One possibility of accounting for why children choose this value first is to assume that actually one value of the null subject parameter is given by UG 'in advance of experience with a given language' (Hyams 1986:8); the task of the child will be to 'reset' the parameter according to the input which he/she receives. Thus, the English child will use null subject sentences in the beginning, since this value is given by UG, it is the initial value of the parameter. Then, because of the input he/she is exposed to, the child will 'reset' the parameter and choose the value which is consistent with the target language.

As we are going to see in 4.3., where various analyses of null subject sentences in child language will be considered, Hyams's solution is not the only one and it is not one without problems. From the point of view of the study of acquisition, one should point out that her approach does not resort to any other theory besides the linguistic one in order to account for parameter fixation. Recall that the approach put forward by Wexler and Manzini resorted to a learning module which interacted with UG. Within the approach put forward by Hyams the initial value of the parameter is given by UG, the child does not have to search for the less marked value. The various possible values of a parameter are fixed in a prescribed order. No learning module is necessary; linguistic theory can also be viewed as a theory of language development.

Among the problems which such an approach faces²² one could mention the one linked to the notion of 'resetting'²³. Allowing for one value to be chosen first and 'corrected' later may lead to learnability problems. If we adopt the view that 'there may also be specific principles of markedness relating various parameters, which need not and may not be fully independent' (Chomsky 1986:146), fixing one parameter for one particular value may trigger fixation of other parameters in a way which is consistent with the combinatorial possibilities. Once certain parameters have been set, it might be impossible for the child to retreat from certain choices, because the PLD will not provide sufficient data for correction. Even when the child has already 'reset' the value of the first parameter, for which there is positive data that lead to correction, for some other parameters, which have been set as a consequence of the fixation of the first parameter in the 'flow', there might be no way back. Obviously, this view relies on the assumption that parameters are not fully independent from one another.

3.2.2.3 Parameter setting as a triggering process

Parameter setting is deterministic and automatic.

We have seen that the process of language acquisition reduces, to a large extent, to the setting of parameters, the 'switches' which can be (automatically) fixed by experience. The idea of language acquisition as parameter setting has been taken to solve the traditional hypothesis-testing models in that the setting process is simple and mechanical and hence can be more uniform across learners. A brief and clear

²² See, among many others, Saleemi (1992) (Chapter 5) or O'Grady (1997) (Chapter 5).

²³ For arguments against resetting of parameters see, among others, Platzack (1996).

explanation of what the automatic flipping of parameter switches on the basis of relevant input could be is the one in Fodor (1999: 366):

What this means, presumably, is that some immediately accessible property of a word sequence is detected by the learning mechanism, and causes a change in the grammar without there being any intervening computation of consequences or evaluation of alternatives.

The advantages of such an account are obvious: no thinking process is required, no rules or generalisations need to be detected, no understanding of the relationship between grammars and the languages they generate is necessary. Evidence does not need weighing. The process is extremely simple, mechanical and, consequently, it can explain why learning is fast and uniform.

However, this view is not without problems. Very often the term of hypothesis or hypothesis formation is still present in studies of language acquisition which adopt the principles and parameters model²⁴. The main idea behind the switch-setting metaphor is the fact that a model of language acquisition should be able to account for the ease with which the child acquires language in the absence of instruction; that acquisition is something that 'happens' to the child, not 'something that the child does'. Hypothesis formation or hypothesis testing refers to a different type or style of learning²⁵ which is neither deterministic nor automatic in the way in which parameter setting is assumed to be.

A second important problem is related to the fact that parameters are not always evident in the input. Some sentences may be ambiguous with respect to parameter value(s), at first sight they can be licensed by more than one combination of parameter values (Fodor 1999).

Also, if one assumes the possibility of parameter re-setting, i.e. that parameter fixation is not a once-and-for-all deterministic fact, at least in some modules, then one must also assume that the child adopts and rejects hypotheses with regard to the appropriate value of the respective parameter. Thus, if a deterministic approach is to be adopted, the hypothesis of resetting parameters is no longer tenable.

The P&P approach tried to reduce the searching space such that parameters could be switched to one of the (probably) two possible values²⁶. The space is further limited by the fact that setting one particular value to one particular parameter may automatically trigger setting one particular value to a cluster of other parameters.

²⁴ See, for example, the following fragment from Chomsky (1986: 55): *In light of the facts of language acquisition, the basic problem is to construct UG so that the class of admissible hypotheses is small, perhaps single-membered.*

²⁵ For a different point of view, according to which parameter setting is defined as hypothesis testing, see Goodluck (1991).

²⁶ This does not imply that one and the same language should choose one single value as a possible option. Chomsky (1992) speculates that Arabic may have both +strong and + weak Tense. If this is true, then this learnability story might need revisiting.

Parameter fixation can be defined as a ‘triggering’²⁷ process, during which the learner can set the value of each parameter possibly only once.²⁸

If one accepts this line of reasoning, the next problem that needs solving is related to what exactly in the input which the child receives can count as a ‘trigger’: a whole sentence, as argued in Gibson and Wexler (1994) or a subpart of a sentence, as argued in Lightfoot (1991, 1998) and Fodor (1998)?

Given the fact that triggering is automatic, the second important question addresses the possibility of mis-triggering and, consequently, of re-setting of parameter values. Can mis-triggering be avoided? If it cannot be avoided, what does the learner do in order to get rid of a ‘wrong’ setting? Within Gibson and Wexler’s learnability model, mis-triggering is allowed: the child may mis-set a parameter and then, on the basis of the PLD, he/she will change the value previously associated with that particular parameter if a certain sentence in the input cannot be syntactically analysed with that particular value. With Fodor (1998), the child can set the parameter value only once.

Let us consider three learning models relying on a P&P approach to the computational system and see what possible answers one might provide to the above questions.

3.3 P&P and Learnability

3.3.1 Cue-based acquisition and degree-0 learnability

Degree-0 Learnability: in order to set parameters, the child has to look for cues in unembedded binding domains.

While ST led to learning models which proposed that children could acquire language on the basis of complex input, the P&P model provided the framework for a theory of learnability whose main idea is that parameters are set by structurally simple data, defined as unembedded binding domains. Hence the name of Degree-0 Learnability (i.e. 0 levels of embedding) (Lightfoot 1989, 1991, 1994). The task of the learner is, within this model, ‘to scan the linguistic environment for cues only in simple syntactic domains’ (Lightfoot 1999:139) where ‘cues’ are defined as ‘some kind of structure, an element of I-language’ (Lightfoot 1999: 139). During the acquisition process the child looks for abstract structures. For example, a cue-based learner of English can set the Specifier-head parameter on exposure to the phrase *John’s hat* (where *John’s* occurs in the Specifier position, preceding the head of the DP, *hat*, when he/she has a partial analysis of the input form as separate words. The learner is assumed to look for such cues only inside structurally simple domains. Let us see then what counts as a structurally simple domain. The hypothesis of degree-0 learnability builds on the assumption that an element’s binding domain is the first CP which contains an accessible SUBJECT²⁹ and that ‘a degree-0 learner sets parameters on the basis of

²⁷ Fodor (1978)(cited in Goodluck 1991) points out that there is a clear difference between triggering and learning. Learning is defined as an externally occasioned change in epistemic state where there is a relation of content between the learner’s representation of the externally occasioning event and the change in epistemic state (for example, if one observes tokens of red London buses one comes to believe that London buses are red; also if one is told that London buses are red one will come to believe that London buses are red). Triggering is an externally occasioned change in epistemic state where there is an arbitrary relationship between the representation of the external event and the change in epistemic state (for example, after receiving a blow on the head, one might come to believe that he/she is the queen of Denmark).

²⁸ As will be shown in 3.3.3, the learning model of Gibson and Wexler (1994) assumes a different view: the learning system randomly chooses one value on the basis of the input and relies on its possibility to correct the value later.

²⁹ An accessible SUBJECT for an element is defined as the AGR of finite clauses or the first c-commanding NP with which that element can be co-indexed without violating any grammatical principles.

data not from embedded clauses, which would obviously be too restrictive, but from unembedded binding domains' (Lightfoot 1994:458). Parameters are set on the basis of data from matrix clauses plus the so-called 'connection points' (i.e. elements which play a part in selectional restrictions, in subcategorisation frames or in linking time reference) with embedded clauses:

- (29) [_{CP} SPEC C [(NP) I...]]

Lightfoot's hypothesis captures the distinction which Emonds (1976) drew between root and structure-preserving transformations or Ross's (1973) Penthouse Principle, which said that operations in embedded clauses represent some subset of operations in matrix clauses. In terms of learnability, this insight translates into the hypothesis that if parameters can be set on the basis of the data provided by unembedded binding domains, then whatever affects embedded domains is a by-product of what affects matrix domains, i.e. the LAD requires only main clauses in the input.

Let us now examine how degree-0 works. Rizzi (1982) argued that one of the parameters which distinguish between languages like Italian, on the one hand, and languages like English, on the other hand, is that in English NP and IP can be bounding nodes whereas in Italian the bounding nodes³⁰ can be NP and CP. This difference could account for why the Italian sentence (30) is correct (*a cui* can move across two instances of IP because IP is not a bounding node in Italian) whereas its English counterpart (31) is incorrect (IP is a bounding node in English):

- (30) *Tuo fratello_i [a cui_i [_{IP} mi domando [_{CP} che storie_j [_{IP} abbiamo raccontato e_j e_i]]], era molto preoccupato.*
 (31) **Your brother_i [to whom [_{IP} I wonder [_{CP} which stories_j [_{IP} they told e_j e_i]]]] was very troubled.*

According to such an analysis, the child would need access to degree-2 data in order to set this parameter. However, Lightfoot (1989, 1991) argues that the parameter can be fixed on the basis of simple triggers. A sentence like (32) can provide evidence that IP is not a bounding node for Italian:

- (32) *Ne_i ho visti [_{IP} [_{NP} molti e_i] corregli incontro]*
 of them (I) saw many run towards him

Ne has moved over NP and IP in a monoclausal structure. The child acquiring Italian can learn that IP is not a bounding node in his/her target language only on the basis of simple data, i.e. from unembedded sentences.

Thus, the child can look for cues in simple, Degree-0 domains. But what kind of structure available in such simple domains can count as a cue?

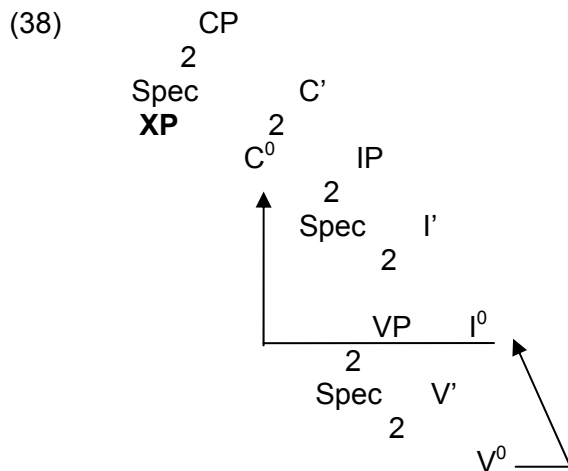
Lightfoot provides an example from the domain of V2. In V2 languages, such as German, Swedish, Danish or Norwegian, the finite verb in the matrix must always occur in second position, whereas the first position is occupied by an arbitrary constituent of type XP:

- (33) [*Pünktlich, korrekt und logisch*] **sind** die Deutschen.

³⁰ A bounding node is a constituent (IP, CP or NP) which represents a boundary for movement. The bounding nodes are subject to parametric variation. The Subjacency condition states that movement cannot cross more than one bounding node.

- Punctual, correct and logical are the Germans
- (34) [*Einige Dinge*] **findet** sie auch positiv.
Some things finds she also positive
- (35) [*Sie*] **arbeitete** als Lehrerin.
She worked as teacher
- (36) [*Hier*] **kann** man Rad fahren.
Here can one bicycle ride-INF
- (37) [*Außer der Bibel*] **hat** sie in ihrem Leben kaum etwas gelesen.
Except for the Bible has she in her life hardly else read

As can be seen in the German examples above, the finite verb is always preceded by an XP constituent: AP in (33), DP-direct object in (34), DP-subject in (35), AdvP in (36) or PP in (37). The finite verb moves to Infl and then to Comp, and the Specifier position of CP hosts an XP constituent:



Whenever the Specifier is lexically filled, the finite verb must move to Comp or else the sentence is ungrammatical:

- (39) **Hier man kann Rad fahren.*
Here one can bicycle ride

How does a child acquiring a V2 language realise that the target language is V2, i.e. that the finite verb in the matrix must always occur in second position? What exactly in the input triggers the correct setting of the V2 parameter? Within a cue-based learnability model, what the child needs to learn is that matrix clauses begin with an arbitrary XP and UG provides the rest of the information: lexical material in the Spec of CP must be licensed by a lexically filled Comp (Lightfoot 1991,1999). The cue is the abstract structure given in (40) below:

- (40) Spec_{CP} [XP]

The cue must be robust, in the sense that it should pass on clear information with respect to the position to which the verb must move. If the XP in the cue is a DP-subject, the cue will not be robust: in this case, there is no clear evidence that the verb must move to Comp, since the DP-subject could also land in Spec of Infl with the finite verb occupying Infl. The input should provide a reasonable number of utterances with initial non-subjects

XPs so that the cue might meet the criterion of robustness.

The child's task is to look for such cues in simple syntactic domains: *children's linguistic development results from their finding certain abstract structures in their chaotic environment, in much the same way that the development of their visual system results from exposure to very specific, abstract, visual stimuli like edges and horizontal lines. This is pretty much what one would expect of selective learning quite generally.* (Lightfoot 1999:19). Data from the acquisition of V2³¹ seem to support the view that children seek for cues in simple domains: Dutch and German children know that the verb can occur in sentence final position before they know how to use embedded clauses.

Lightfoot provides further arguments in favour of his hypothesis from the area of language change and from the domain of creole languages, showing in what way his learnability model can also account for the change and development of languages. A very important assumption which lies at the core of the model is that the child is not seeking to match the input. According to Lightfoot, any input-matching model of learnability cannot account for language change or for the creation of creoles, where the child creates an output grammar different from the one provided by the speech of the elders. For example, the Guyanese creole language Berbice Dutch is an SVO language. The two languages on which it is based are Dutch and the Kwa language of Eastern Ijo, which are both underlyingly SOV, order which is surfaced as such in embedded clauses but almost never in matrix clauses. This empirical fact, argues Lightfoot, points to the fact that the Guyanese children relied on matrix clauses when setting the VO parameter. If they had relied on embedded domains, Berbice Dutch would be an SOV language, not an SVO one.

The Degree-0 hypothesis is both restrictive and quite radical³². However, if one assumes a P&P approach to language acquisition, this hypothesis is in the spirit of the model. In spite of the problems such a learnability theory might raise, it has pointed out that one should look for principled constraints on the type of data on which the LAD can depend.

A cue-based model of acquisition can also account for the possibility of language change. If one adopted the view that the child is looking for grammars which necessarily match the input', one could not account for why certain parameter values changed in the history of particular languages.

3.3.2 The structural templates hypothesis

Fodor (1998) develops a similar learning model. She defines triggers as 'small structural templates that are innate, are stored by the language faculty, and constitute the parametric options offered by UG for languages to make use of if they choose to' (p.19). The task of the learner is to detect which of these templates (or 'treelets') are for the target grammar. The learning system which she creates is able to avoid mis-triggering due to ungrammatical or unambiguous input, i.e. a system which is endowed with the skill of detecting ambiguous/unambiguous input and respond differently to each of the two types. The learning model she adopts is the one in which parameters can only be set once. If the input is ambiguous, the learning system simply 'waits' for unambiguous input to set parameters.

³¹ For a different analysis of the acquisition of V2 (in German) see Weissenborn (1999).

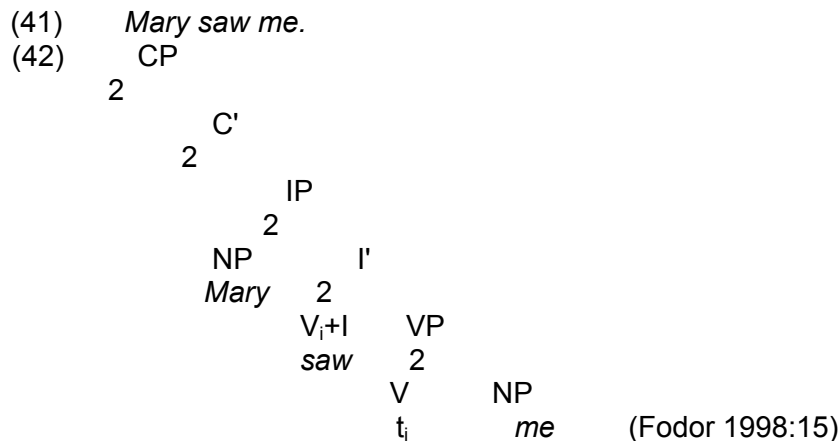
³² For commentaries on Degree-0 learnability, see the Open Peer Commentary section in Lightfoot (1989). Also, for arguments in favour of the need of embedded domains for parameter setting, see Roeper and Weissenborn (1990).

Within her model triggers are allowed to include structure, they are 'structural templates', 'syntactic subtrees' (p.18):

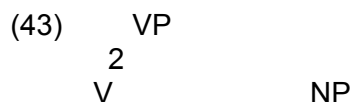
Each parameter value is thus associated with its own structural (or in the limiting case, featural) signature, that is, with whatever constitutes its essential contribution to the sentence structures that are licensed by grammars which have that value. (Fodor 1998:17)

One problem still remains to be solved: how does the learner (come to) know which subtree, which treelet can count as an appropriate trigger? The answer is more difficult to find than it might look at first sight since with Fodor many parameters cannot be set on the basis of the 'surface' input. Many sentences may be ambiguous in this respect. This suggests that the child has to rely on an analysis of the underlying structure. Sentences have to be assigned a full structural description of all levels of derivation before the child can detect the cue which will trigger the appropriate setting of the parameter value related to that cue. The claim is that treelets are underlying syntactic structures which act as global triggers. For example, a sentence like (41), with the structure in (42), is an unambiguous trigger for the parameter value Verb-Object (VO):

The cues to parameter fixation are underlying syntactic structures.



But the VO value of the parameter is not given by the surface word order, but by the fact that the direct object is to the right of the trace of the verb, i.e. by the underlying order of constituents. Which actually means that the treelet acting as a trigger is the one in (42); what is relevant for the setting of this particular parameter, is not the whole sentence, but only the structural sub-tree of the whole clause given in (43):



Learning is defined as incremental, in the sense that only one parameter can be set per input sentence, even when the sentence provides an unambiguous input for several parameters. Treelets serve both as the parameter value and as its trigger. More significantly, these triggers are global; they occur in every language which evinces that particular parameter value.

How does the learner recognise a trigger on the basis of the input which he/she receives? As a trigger, the treelet must be identified in the structure assigned, via parsing, to the input sentence. The learner tries every grammar out on this input sentence in order to see if more than one of them is successful. The sentence counts as unambiguous only

when one single grammar is successful. Fodor identifies this recognition problem with a perceptual problem and reduces it to the ideal parsing test in which the learner parses the input sentence only once and knows which structural triggers are contained in the input. The child knows the structural triggers because they are part of UG³³. Thus, within this learning model, the grammar for a particular language consists of: (i) the principles of UG; (ii) a lexicon; (iii) universal structural resources, as for example, X-bar format; (iv) a set of parameter values/structural triggers provided by UG. With this grammar, the learner parses the input sentences. If only one analysis is found, the input is unambiguous and the learner fixes the parametric values. If more than one analysis can be found, the input is ambiguous and hence it is safe for the learner to wait for an unambiguous input.

Such a view differs from the one put forth in Chomsky (1995), for example, where parameter values are metaphorically associated with the different positions of a switch. Within Fodor's model, both values of a parameter are provided by the grammar with which the learner parses the input. But only one value can be used during a parse.

One advantage of this model is that children are no longer attributed psychological processes different from those of adults. The trigger structures they need in order to parse the input are the ones which adults use in parsing the same input. Also, we avoid the 'resetting'- of- parameter- story, which is not without problems. At the same time, it is more in the spirit of the 'deterministic' parameter fixation process adopted within a P&P approach to acquisition and, by putting forth the hypothesis that triggers are 'sub-trees' and not whole sentences it can account in a more elegant way for the acquisition process in the absence of a rich, non-deficient input.

3.3.3 The Triggering Learning Algorithm

Unlike Lightfoot or Fodor, Gibson and Wexler (1994) take whole sentences as possible triggers: 'there are sentences in the child's experience that point directly to the correct settings of the parameters. [...] for any setting of a non-subset parameter, there is a sentence that is grammatical under that setting but not under any other'³⁴ (p. 408). The parameter space is assumed to unambiguously signal the value of one single parameter, i.e. small changes are preferred to larger ones. They posit the so-called Triggering Learning Algorithm under which the child uses his/her current grammar to syntactically analyse sentences provided by the input and changes the previous hypothesis about this current grammar (for example, the value of a certain parameter) only when the sentence cannot be syntactically analysed on the basis of the previous hypothesis (the TLA is error-driven). The algorithm analyses the next piece of input. If the sentence can be accepted as 'grammatical' using the value already assigned to that given parameter (i.e. using the current grammar), nothing happens. But if the sentence cannot be parsed, the algorithm will change the value of the parameter.

The child has to analyse the whole sentence in order to fix the value of a parameter.

A parameter is (more often than not) randomly selected (if it does not have a default initial value) to have its value changed and the new value is adopted (at least for a while) only when it allows the syntactic analysis of the new sentence. The new value is not necessarily the correct one: it only allows the analysis of the current sentence and it may have to be changed again. However, the child will finally converge on the correct grammar: it is assumed that there is always at least one trigger for an incorrectly fixed

³³ In this, Fodor's view is along the line put forth by Dresher and Kaye (1990) or Dresher (1998), according to which UG provides both the set of parameters and a cue for each parameter.

³⁴ This reminds of Morgan's Degree-1 learnability model, since no embedded sentences are considered possible triggers.

parameter value. In the process, the learner relies only on positive evidence. Memory of previous parameter settings is not required.

One can see that, on such an account, parameters can be set and re-set, in a step-by-step process. The learner changes one parameter at a time, since, as has already been said, small changes are preferred.

Within Gibson and Wexler's learnability model, triggers fall into two classes: global and local. A global trigger for the value v of a parameter is defined as a sentence that is grammatical if and only if the value of the parameter is v , no matter what the values for other parameters are (p. 409). For example, the pattern Subject Verb works as a global trigger for the value Spec-first of the specifier-head ordering parameter. A child acquiring English, upon hearing the ordering Subject Verb, will end up hypothesising a Spec-first value in order to be able to syntactically analyse the current sentence. So far, no value for the complement-head parameter has been selected, since a trigger of the type Subject Verb allows for both comp-final (SVO) and comp-first (SOV) orderings. The value of the Specifier-head parameter has been selected without taking into account the comp-head ordering. Since English is an SVO language, how does the child converge on the correct grammar? The local trigger, a sentence of the type Subject Verb Object, which is correct only if the value for the comp-head parameter is comp-final, will allow the child to fix the correct value of the comp-head parameter and hence converge on the correct English grammar.

However, there are certain situations for which there exist no triggering data, be they global or local.³⁵ The learning algorithm is 'trapped'. This so-called *local maxima* problem can be, nevertheless, avoided because (i) some parameters have default values and (ii) some parameter values may mature. An example of such a situation is the one related to the V2 domain in Dutch and German. What can the TLA do for the German or Dutch child to help him/her finally fix the correct parameter value? It is hypothesised that some parameters (one or two) receive a default initial value from the set of possible values while others are initially unset. The child begins with default initial values being allowed to change only a subset of parameters for an initial period of time. The other parameter(s) will be considered at a later stage. Thus, in the case of V2 languages, it is assumed that there is a default value for the V2 parameter, -V2. The child will only try to fix the values of the specifier-head and complement-head parameters: if the hypothesised grammar does not allow the analysis of the current sentence, the value of the parameter will be changed. It is only after the values of these two parameters have been set that the child considers the alternative value for the V2 parameter.

Gibson and Wexler's model actually implies that the learner hypothesises that there are two types of parameters: parameters whose value is not fixed and parameters which can have a default value. Moreover, the default value is assumed to be different from the correct value of that particular parameter in the target language. This further implies that the child simply knows he/she does not have to consider some parameter(s) for a certain amount of time. The hypothesis from which the child starts with respect to some parameters must be false. The model also stipulates that there is a fixed parameter ordering: the child has to consider the parameters whose value is not pre-specified first and only later (maturation?) to consider the parameter(s) with a default setting. It does not, however, say anything with respect to the ordering of the setting within the class of unspecified-value-parameters³⁶.

³⁵ Actually, in the end the authors abandon the idea of global and local triggers in favour of the so-called deductive triggers.

³⁶ For a more detailed discussion on how the model works, as well as on its advantages and disadvantages, see Dresher (1998) or Berwick and Niyogi (1996).

4. Acquisition from a Minimalist Perspective

4.1 *New focus of inquiry*

The Minimalist Program (MP) (Chomsky 1993, 1995, 1998, 1999) builds on assumptions already present in the P&P, but with a shift in focus. The main area of inquiry regards the optimality of design of the language faculty. A genuine explanation of why language is the way it is represents the goal of minimalist inquiry. The program is not really what one would have expected in terms of descriptive strength. But it certainly represents a gain in terms of a deeper understanding or at least deeper inquiry into the nature of the computational system of language.

The tension between descriptive and explanatory adequacy is this time being solved with an obvious bias towards the explanatory part. The puzzling link between the derivational approach to language and the condition of explanatory adequacy, on the one hand and the nature of the mind/brain on the other hand, has never before been as clearly and strongly assumed. The Minimalist Program is not, actually, a fully developed theoretical model which readily allows for the technical description of empirical data. It is a 'program' which provides a framework of inquiry into the study of language³⁷. The Minimalist Program as presented in Chomsky (1998, 1999) already revisits ideas presented in Chomsky (1995) and Chomsky (1993). But the new solutions are all in the spirit of the program, they do not represent a radical shift from the core as they do not represent a radical shift from earlier generative assumptions (especially the ones of P&P), which find a better explanation or a better formulation or which are now questioned from a perspective enriched with the results of generative inquiry. Language is still viewed as a component of the human mind/brain, a faculty with which humans are endowed from birth, uniform for the species. Linguistic development is defined, as in previous models, as the change from an initial state S_0 ('constituted of invariant principles with options restricted to functional elements and general properties of the lexicon' Chomsky 1995: 170) to a final state S_f , an idealised I(nternal)-language, a complete specification of parametric option, on the basis of experience (PLD). UG provides the invariant principles of S_0 as well as the parameters which represent the range of permissible variation. The approach to language is, as always, derivational, and linguistic expressions are defined as a sequence of representations along the derivational process. Language is seen as consisting of a lexicon and a computational system. What is then the contribution of the Minimalist Program to the study of language?

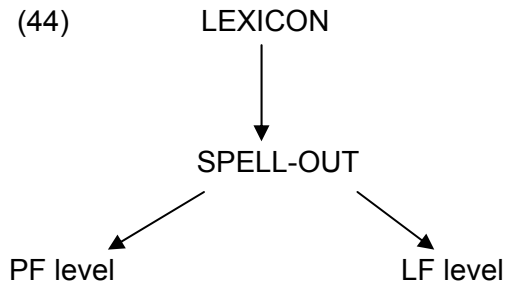
In what follows, let us briefly present how the program views the lexicon and the computational system as well as the relation between them and how this may shed new light on old questions, with a focus on the questions related to acquisition.

³⁷ 'At the methodological level, the program has certain heuristic and therapeutic value. It brings to light what might be fundamental problems, where empirical evidence and minimalist expectations conflict. And it encourages us to distinguish genuine explanations from "engineering solutions".' (Chomsky 1998: 5).

4.2 The Minimalist Program in a Nutshell

4.2.1 The general organisation of grammar

The general organisation of grammar assumed within a minimalist approach is the one in (44):



MP: only the external levels (PF and LF) have been pre-served. Linguistic objects must satisfy conditions imposed by these two levels.

As can be noticed, the assumed organisation is simpler than the one adopted within the P&P model. The only available levels are the external ones: the level of Phonological Form (the articulatory-perceptual interface) and the level of Logical Form (the conceptual-intentional interface, the place where linguistic expressions interface with the cognitive systems). D-Structure and S-Structure have been eliminated. Within a ST approach, Deep Structure represented the only input to the semantic component, with surface structures playing no part in the semantic interpretation of linguistic representations. In a P&P model, D-Structure represented the internal interface where the relation between the lexicon and the computational system was expressed. The information at D-structure was no longer assumed to be directly sent to the semantic component. It was mapped to the level of S-Structure and then to LF. Essentially, the D-Structure information as such was not directly accessible to the performance systems. Within the minimalist model, each derivation has to meet the interface conditions: those of the PF level and those of the LF level. Some parts of the computational system are relevant for PF, other parts are relevant for LF. A linguistic object is defined as a pair (a, b), a formal representation of sound and meaning, satisfying the PF and the LF conditions. When these are met, the derivation is convergent. When they are not met, the interpretation is deviant. As can be seen, the role of the interface levels, i.e. the point of interaction between language and the sensorimotor systems, on the one hand, and language and modes of thought and understanding, on the other hand, gains in significance. On such a view, there is no justification for any additional level. Observed properties of language are accounted for in terms of legibility conditions imposed by the two available visible levels and by properties of the computational system of language.

What are the consequences of this new picture for the computation? On P&P assumptions, the information in the lexicon was projected to D-Structure via the Projection Principle and constrained by the Theta-Criterion, with the tension regarding the primacy of syntax vs. semantics already mentioned. With no D-Structure available, there is no justification for the two principles to be postulated and the tension disappears³⁸. Postulating only two interface levels, the external ones, as well as preferring operations required by these interfaces represents an obvious gain in the direction of explanatory adequacy which requires that 'parameters be easily detectable for language acquisition' (Chomsky 1999:2). Also, importantly, the complexity of computation is reduced.

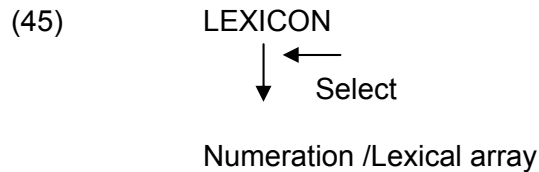
³⁸ But see Uriagereka (1999) for the defence of D-Structure within the Minimalist Program.

4.2.2. From the Lexicon to PF and LF

4.2.2.1 Remarks on projections

Let us see now how the information from the lexicon is taken over by the computational component within a minimalist model. An array of lexical items is selected from the Lexicon via an operation *Select*. The array of lexical items selected from the Lexicon is called Numeration and it contains, in simple terms, the entities from which syntactic objects, such as sentences, will be formed:

Syntactic objects are created via Merge and Move.



A Numeration consists of pairs of the type (LI, i) , where LI stands for a linguistic item and i for its index, i.e. for the number of times the item has been selected in the Numeration. Its psychological counterpart could be defined as 'what the speaker has in mind'.

Projections are created via Merge, externally determined.

The computational system, which can access only the Numeration, not the whole lexicon, maps the Numeration into a syntactic object (LF and PF can interpret only single objects), via permissible operations which recursively construct syntactic objects, until all indices are reduced to zero. The core operations which apply are *Merge* and *Move*.

The minimalist definition of the lexicon does not radically differ from the one adopted in previous studies. However, it is worth pointing out that the inventory of functional categories available in the lexicon is reduced:

It is clear that the lexicon contains substantive elements (nouns, verbs) with their idiosyncratic properties. And it is at least reasonably clear that it contains some functional categories: complementizer (C), for example. But the situation is more obscure in the case of other possible functional categories, in particular T, Agr, specific phi-features, a Case category K, and so on... (Chomsky 1995:240).

This suggests that functional categories may actually fall into at least two different classes, with different properties, whose acquisition depends on the amount of computation and/or lexical learning involved in the process.

One more thing worth pointing out is that the MP focuses more than the P&P model on the idea that all categories are to be defined as sets or bundles of features³⁹.

In earlier versions of the MP (1993) the information stored in the lexicon simply projects to an X-bar structure, i.e. the format of X-bar is still assumed to be provided by UG, just as in P&P. In Chomsky (1995, 1998, 1999) projections are assumed to be determined by the relational properties of the categories which constitute them and are

³⁹ One should stress, however, that this idea goes back to Chomsky (1970) and that it underlies X-bar theory: 'In the earliest work in generative grammar it was assumed that the elements of the underlying base grammar are formatives and categories; each category corresponds to a class of strings of formatives. [...] it was soon found necessary to depart from this assumption in the case of lexical categories [...]. We might just as well eliminate the distinction of feature and category, and regard all symbols of the grammar as sets of features.' (Chomsky 1970:48).

created via Merge, with standard X-bar theory being thus largely eliminated. Projections are the result of a simple operation, Merge.

4.2.2.2 Merge

Merge is the simplest operation which takes a pair of syntactic objects, forming a complex syntactic unit out of the ones provided by the Numeration⁴⁰ or already constructed in the derivation. Suppose we have two objects *a* and *b*. In earlier versions of the MP, if *Merge* applies to these two objects, a new larger object will be created: $\{c, \{a, b\}\}$, where *c* stands for the label of the new object, determined derivationally: if *a* projects and is the head of the new object, its label will be *a*: $\{a, \{a, b\}\}$. If *b* projects, the label of the new object will be *b*: $\{b, \{a, b\}\}$. Merge (*a, b*) is thus asymmetric, allowing either *a* (46a) or *b* (46b) to project:

- | | |
|---|---|
| <p>(46) a. a
 2
 a b</p> | <p>(46) b. b
 2
 b a</p> |
|---|---|

The ‘wrong’ option is filtered out at LF, by various principles. For example, if a verb like *eat*, and a direct object DP, *chocolate*, Merge, the option with a DP label (47b) is filtered out by theta theory, with (47a) being the only possibility:

- | | |
|--|---|
| <p>(47) a. VP
 2
 eat chocolate</p> | <p>(47) b. *DP
 2
 eat chocolate</p> |
|--|---|

The X-bar format is thus radically simplified, the idea of binary structure is the only one that has been preserved.

Why is it desirable to replace the X-bar format with *Merge*? What does it actually mean for the computational system? On these assumptions, items from the lexicon project not because of some postulated format rules, which are given on grounds of virtual conceptual necessity, but because of conditions imposed by the interface. Only single objects can be interpreted by LF and PF; smaller objects must hence merge into a larger one which can be interpreted. *Merge* is externally determined. Recall that when two objects merge, either can be a head and project. Thus, the headedness property is accounted for by the *Merge* operation and hence derivative from external conditions.

The child will have to learn the properties of the items stored in the lexicon and set the correct value to parameters. No format rules have to be assumed anymore. Lexical and functional items are projected via a simple, binary operation, *Merge*. In terms of computational cost, this is the least costly operation. If one assumes that computational complexity matters for linguistic development then one can also assume that *Merge* should be the only operation present in early child syntax during very early stages. This hypothesis is borne out by data. *Merge* receives direct representation in early child syntax which it can account for and from which it can, at the same time, receive strong motivation. The existence of *Merge* allows us to account for early lexical stages in acquisition, when children adjoin lexical items between which there is no unique relation in adult grammar:

- (47) a. *more car/ more cereal/more read/more hot/ more walk*

⁴⁰ See Collins (1997) or Johnson and Lappin (1997) for arguments against the use of numerations from the perspective of performance theory.

- b. *no bed/no home/no wet/no high/no fix/no plug* (Powers 1998)
 c. *are you put this on me/are you help me/are this is broke/are you sneezed* (Roeper 1994)
 d. *toto auch/ich auch/hauschen auch* (Tracy, Penner and Weissenborn 1993)

Items such as *more* or *no* represent, at this early stage, heads of Merge projections:

- (48) a. $\begin{array}{c} 2 \\ \text{more} \end{array} \quad X$ (48) b. $\begin{array}{c} 2 \\ \text{no} \end{array} \quad X$

They differ from other entities: they are lexically specific (Roeper 1996) and semantically stable (Powers 1998) and their projection is not to be found in adult grammar. Such phrase markers exist only in child syntax. Powers (1998) proposes that these elements behave like semi-lexical items defined by the following properties:

- (i) they have the same semantic features as target lexical items;
- (ii) they have different (or none of the) syntactic features than (of) target lexical items;
- (iii) seemingly, they have the same distribution as truly grammatical functors.

The operation of *Merge* makes the existence of such representations possible. Once the Numeration has been selected, elements are syntactically connected as Spec-head (49a) or Head-complement projections (49b), before the emergence of the more complex specifier-head-complement structures (49c)⁴¹ (Roeper 1996):

- (49) a. $\begin{array}{c} 2 \\ \text{Specifier} \quad \text{head} \\ \text{Mommy} \quad \text{sock} \end{array} \quad (\text{used for Mommy's sock})$

- b. $\begin{array}{c} 2 \\ \text{head} \quad \text{complement} \\ \text{more} \quad \text{car} \end{array}$

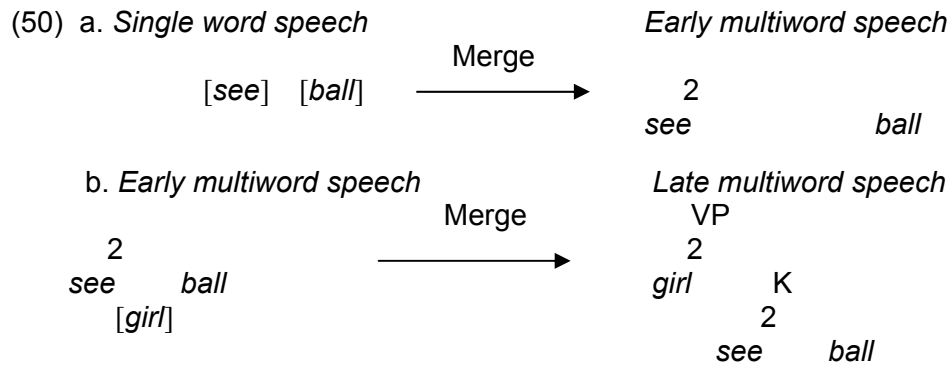
- c. $\begin{array}{c} \diagup \quad \diagdown \\ \text{Specifier} \quad \text{Head} \quad \text{complement} \end{array}$

The acquisition sequence = successive applications of *Merge*.

An approach to syntactic development in terms of *Merge* can elegantly capture the acquisition sequence from the single-word speech to later multiword speech, which can be now defined as successive applications of *Merge*:

[...] *the operation Merge applies and reapplies to the syntactic objects (the phrase markers already constructed) yielding longer and more complex structures.* (Powers 1998: 4)

⁴¹ Notice that if this assumption is true, its implication will be that in early child syntax no distinction is made between the complementizer and the specifier positions. In the 1993 version of the MP, the complement and the specifier represent different domains, the checking and the internal domains of the head, and the specifier position is associated with *Move*, an operation which is not available at an early stage according to this developmental account.



The existence of these *Merge* structures, ‘micro-steps of acquisition’ (Roeper 1994:14), in early child speech, whose sequence reflects the sequence of syntactic representations, provides evidence for *Merge* as a core property of the computational system, especially if one adopts the view that *the grammar constructed by the child is a derivational one, and that this is later converted into alternative formats [...]. This would mean that the study of acquisition [...] would give a unique purchase of grammar in the derivational mode* (Lebeaux 1988).

A few remarks are in order here with respect to the implications of the earlier definition of *Merge* (Chomsky 1993, 1995) for the computational complexity involved in the process. On the one hand, *Merge* is assumed to be the simplest, the least costly operation. On the other hand, providing the appropriate label to the associative object created via *Merge* was said to be decided on only at LF. This implies that one has to ‘look ahead’ in order to avoid the ‘wrong’ projection. Or, looking ahead, in its turn, implies substantial computational complexity. We are thus faced with an operation which is defined as the least costly but which involves, at the same time, computational complexity. The implications for acquisition are obvious. In order to avoid this ‘puzzle’, in Chomsky (1998) it is proposed that the label of the new syntactic object created via *Merge* is determined by the selectional properties of the merging elements. For example, a verb and its direct object DP merge to satisfy the requirements of the verb which, being the *selector*, will determine the label of the new object:

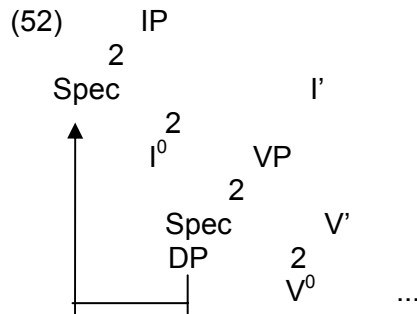
(51) {eat} {chocolate} → {eat{eat, chocolate}}

The wrong label is no longer detected as late as LF and ‘look ahead’ is eliminated since only one single label is possible. *Merge* is indeed the least costly operation. But, if it is now defined as relying on the selectional properties of the items in the Numeration, what would the implications for acquisition be? How can one account for those early two word structures which are made of two items adjoined according to relations different from the ones available in adult syntax (see 47)? There seems to be no selection involved in *more car*, for example. In Chomsky (1998) a distinction is made between *set-Merge* and *pair-Merge*. The former operation can apply only to two elements which stand in a kind of selectional relation whereas the latter does not involve such a relation (for example, the merger of an adjective and a NP). Early grammars may use only *pair-Merge*, possibly because the selectional properties of items have not been acquired yet.

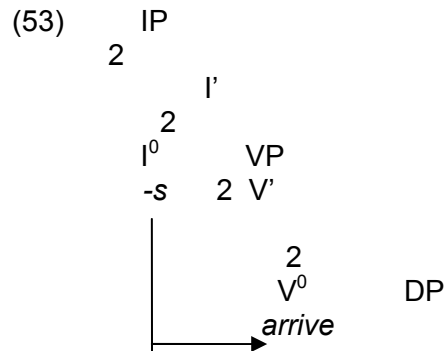
4.2.2.3 Move

The second important operation is *Move*. Within a P&P model, any syntactic object could be moved to a target position leaving behind a trace with which it formed a

chain. Constraints on movement were derivative from various principles. For example, under the VP-Internal Hypothesis the subject was assumed to be generated in the Spec position of the VP from which it then had to move (in languages like English) to satisfy the Extended Projection Principle⁴²:



Crucially, items were taken from the lexicon uninflected and inflectional morphemes were assumed to head projections on a par with lexical items. In (53) below the affix *-s* is assumed to lower and adjoin to the bare, uninflected lexical verb. Then PF rules will interpret the complex *arrive* + *-s* as a single inflected phonological word.



Within MP, Move is still defined in the same way: the operation via which an item moves to a targeted position leaving behind a trace with which it forms a trace. The crucial difference is that such operations are now assumed to be driven by morphological necessity, by features associated with functional categories such as Tense, Agreement or Determiner. Items are drawn from the lexicon fully inflected. *Arrives*, for example, has all its inflectional features (Tense and Agreement) in the lexicon and enters the derivation as such. Affixes are associated with features which have to be checked against the appropriate inflectional projection, i.e. they have to agree or to match the (abstract) features represented in the functional heads. Abstract features seem to be more relevant for the computational system than the overt morphology itself.

Overt movement is seen as feature-driven movement, in an attempt at checking the so-called strong (uninterpretable) features which are not allowed to survive at LF. The dislocation property of language is thus associated with uninterpretable features. Weak (interpretable) features can procrastinate, they can wait until LF, where they trigger covert movement. PF can only 'see' the lexical item, features are not visible at this interface level. Once they have been checked, they disappear and the item can Spell-Out, since interface requirements are met. If a feature remains unchecked, the derivation crashes. Notice that *Move* is driven by morphological features, which differ from one language to

⁴² The Extended Projection Principle (EPP) requires that sentences must have subjects.

another. Movement is not required by some general principle, but by language-specific properties which can elegantly account for the different word orders available cross-linguistically in terms of morphological variation. Word order is visible in the linguistic input; being determined by morphological features, it can help the child to determine the properties of grammatical formatives.

Move is subject to conditions of economy: it must choose the shortest possible route, either by choosing a smaller number of rule applications (i.e. the shortest derivation) or the shortest possible movement (the so-called Minimal Link Condition).

Economy of derivation is assumed to require a minimum of transformational activity. *Move*, as has been shown above, being a composite of sub-operations, is computationally more complex than *Merge*. This is why it has to apply as late as possible (as *Last Resort*) in the derivation: an element can move only when it really has to move, otherwise it must stay *in situ*.

Such an account of movement is not without problems. On the one hand, the distinction between covert and overt movement is reduced to the distinction between strong and weak features which means, at the conceptual level, that we have to stipulate that movement is driven by *the feature of a feature* (Solà 1996, Chomsky 1998). Also, Procrastinate induces 'look-ahead', i.e. one has to look ahead in the derivation in order to check whether Procrastinate is justified, which is undesirable at least in terms of computational complexity. This approach to movement is also challenged by instances of movement which is not triggered by the need to check uninterpretable features. One such example is EPP (Extended Projection Principle). In the 1995 version, EPP was assumed to be triggered by an uninterpretable feature on Tense. But what EPP actually means is that a DP must be merged next to the predicate, in sentence initial position. Whether the feature on Tense is /is not interpretable is irrelevant (Chomsky 1998, 1999). There are also instances of feature checking without movement, such as the agreement relation between an expletive and the predicate. All these problems led to the abandoning of the distinction between strong and weak features and the intuition that different movements apply at different stages in the derivation (captured in early minimalism in the distinction between movement in overt syntax and covert movement at LF) is now captured by the definition of the derivation as a sequence of phases, with Spell-Out being cyclic (Uriagereka 1997, Chomsky 1998, 1999).

What are the implications of all these facts related to movement for acquisition?

Firstly, in terms of complexity, we have seen that *Move* requires a more complex computation. If we adopt the view that early grammars can only cope with simple computations, then we expect *Move* and phases comprising movement to be available only at later stages of linguistic development⁴³.

Secondly, the Multiple Spell-Out approach to derivation may represent an appropriate framework which can account both for adult and child syntax. Assuming that computational complexity matters for cognitive development, the main difference between child and adult language will be related to the speaker's capacity of dealing with complex operations, i.e. it would reduce to processing capability. Early two-word utterances, for example, could be defined as the outcome of a 'simple' phase, which does not require any costly or heavy operations. Syntactic development would thus be tied to the child's growing processing capability.

On minimalist assumptions the initial state of the language faculty is defined as comprising an array of (invariant) properties and operations which fall into two classes: assembly operations (which create lexical items) and computational operations (responsible for the formation of more complex expressions out of lexical items formed

Move is driven by morphological necessity; the child has to set the correct options with respect to the functional categories in the target language on the basis of the word order visible in the input.

⁴³ See 4.1 for arguments that there is movement in early child grammar.

The language faculty comprises an array of properties, computational operations and assembly operations.

by assembly operations). This suggests that the child has to notice and learn whatever falls under ‘variation’ which includes, besides functional categories (or non-substantive parts of the lexicon), PF options, lexical arbitrariness (association of concepts and their phonological matrices in the traditional sense), and general properties of lexical items. The child has to set the correct options with respect to the properties (the features) of lexical and functional items on the basis of what is visible in the input. This is possible because ‘variety is restricted to easily detectable properties of utterances’ (Chomsky 1992:2).

SUMMARY

The main goal of this chapter has been to show that the core cluster of assumptions with regard to language acquisition within the generative approach has remained unchanged throughout its various models: the Standard Theory model, the Principles and Parameters model and the Minimalist Program. The child is endowed with the ability to acquire language, with the faculty of language, which undergoes changes under the influence of the linguistic input and possibly of maturation. The input-output relation is mediated by a device, call it the Language Acquisition Device (LAD), which ‘filters’ the linguistic input the child is exposed to. It consists of general principles which guide the process of acquisition, restricting the set of possible grammars. And it is exactly the description of this LAD together with the assumptions with respect to the language computational system which have changed from one model to another, in an attempt to meet explanatory adequacy.

	ST	P & P	MP
LAD	Substantive and formal universals + evaluation measure	Principles and (un-set/pre-set) parameters	An array of (invariant) properties and computational and assembly operations

The theoretical assumptions with respect to Universal Grammar as well as with respect to the operations of the computational system led to various learnability accounts:

- A. On **Standard Theory** assumptions, the child’s task is defined as constructing the transformational component on the basis of the input. The question of what exactly in the input was necessary for the child to be able to detect the appropriate transformations received two main answers:
 - (i) the child needs to consider phrases which contain no more than two embeddings (Degree-2 Learnability);
 - (ii) the child needs to consider phrases which do not contain more than one embedding (Degree-1 Learnability).
- B. On **Principles and parameters** assumptions, the child’s main task in the domain of syntax is to set appropriate values to parameters on the basis of the linguistic input. The question is, this time, what exactly in the input can help the child to set this value:
 - (i) the child has to look for cues in unembedded binding domains (Degree-0 Learnability);
 - (ii) the structural templates which trigger parameter fixation are innate and stored in the language faculty. The child must detect the

- underlying syntactic structure (a treelet) associated with each parameter value appropriate in the target language (the structural templates hypothesis);
- (iii) the child has to analyse whole sentences in order to fix the value of a parameter (only one at a time), using previous hypotheses about the target grammar. If the sentence cannot be accepted as grammatical on the basis of a previous hypothesis, a new hypothesis (i.e. a new parameter value) will be adopted (the Triggering Learning Algorithm).

C. Since **Minimalism** is still a research program, there is no learnability account grounded in this model.

The second goal of this chapter has been to show how data from language acquisition as well as the need to account for the process of linguistic development have led to significant changes in the generative model, which has always had as a major task the explanation of acquisition.

Further Reading

Advanced: If you want to get a comprehensive view of the generative models discussed in this chapter and “read” them with your own mind, you must, by all means, go to the very source: Chomsky (1965), Chomsky (1981) and Chomsky (1995). For an unusual but extremely captivating presentation of minimalism (and much more), read Uriagereka (1998). Atkinson (1992) is an excellent discussion on how the P&P model deals with acquisition.

Textbooks: If you are a textbook person and a beginner, try Cook (1988), Radford (1988) and Haegeman (1991). If you already have some knowledge of generative grammar and want to know more, read Lasnik and Uriagereka (1988).

Focussed: If you are interested in matters of learnability models grounded in the generative model, read the very studies presented in this chapter. If you want to find out more about how acquisition could affect language change, Lightfoot (1999) is a good choice.
