On how Learning Mechanisms shape Natural Languages

Adriana Fasanella Seligrat

Universitat Autònoma de Barcelona

Departament de Filologia Catalana

Centre de Lingüística Teòrica

Supervised by Dr. M. Carme Picallo Soler

A doctoral dissertation submitted for the degree of

PhilosophiæDoctor in Cognitive Science and Language

2014 April

Adriana Fasanella Seligrat Dr. M. Carme Picallo Soler

Abstract

The goal of this thesis is to argue, on theoretical and empirical grounds, that mechanisms in charge of language acquisition define both synchronic and diachronic patterns of linguistic realisation found across natural languages. This aim leads us to explore, then, how Greenberg's problem is derivative from Plato's problem. An approach to the logical problem of language acquisition in terms of a modified discovery procedure is proposed, which, based on learnability considerations, anchors the acquisition of abstract properties on perceptible ones. Within this enterprise, a learning constraint and a mechanism of data analysis, both instances of Third Factor principles, are argued to be at work in the growth of language in the individual. From the effects that the learning constraint and the mechanism of data analysis have on the resulting acquired I-languages, some patterns of synchronic and diachronic variation are derived.

Locura tal vez, y locura grande, querer penetrar en el misterio de ultratumba; locura querer sobreponer nuestras imaginaciones, preñadas de contradicción íntima, por encima de lo que una sana razón nos dicta. Y una sana razón nos dice que no se debe fundar nada sin cimientos, y que es labor, más que ociosa, destructiva, la de llenar con fantasías el hueco de lo desconocido. Y sin embargo...

Del sentimiento trágico de la vida en los hombres y en los pueblos
Miguel de Unamuno

Acknowledgements

I thank my parents, José Fasanella Penella and Felisa Seligrat Martínez, and my sister, Assumpta Fasanella Seligrat, for a life full of joy, happiness and laughter. I have learned all important things at home; I cannot express how much I admire the three of them.

I thank Jordi Masip because he always was, and still is, an endless source of inspiration.

***

I thank my supervisor, Carme Picallo, for her constant support and for having shared her linguistic knowledge and passion with me during these four years. Carme has always helped me and encouraged me respecting my own decisions and points of view; for that reason and many more, I feel more than grateful to her.

I have done a PhD in theoretical linguistics because I attended an undergraduate course on *Syntactic Theory* given by Joana Rosselló, and I have done a PhD about the relationship between acquisition and linguistic variation because she first suggested 'parameters' as a research topic (she was also the first in mentioning the name of my future husband, but that is another story). I thank Joana for all that, and for her advice since I was an undergraduate student.

I thank all the committee members, their acceptance is a great honour for me: Dr. Acedo-Matellán, Dr. Bartra, Dr. Fernández Soriano, Dr. Roberts and Dr. Rosselló.

I thank the Centre de Lingüística Teòrica (CLT) for having admitted me as a PhD student. CLT provides one of the best atmospheres one could be involved in to write a dissertation. I thank all its members, specially Víctor Acedo-Matellán, Anna Bartra, Eulàlia Bonet, Josep Maria Brucart, Maria Teresa Espinal, Ángel Gallego, Anna Gavarró, Maria Lluïsa Hernanz, Joan Mascaró, Jaume Mateu, Gemma Rigau and Jaume Solà for their help and discussion on some topics. Particularly, I want to thank Ángel Gallego for

his support since the beginning and for having taken me into account for always interesting projects.

Undoubtedly, the best part of having done a PhD has been my colleagues during the undergraduate courses, the master and at the CLT: Eduard Artès, Teresa Blasco, Isabel Castro, Elena Ciutescu, Javi Fernández, Elías Gallardo, Anna Gibert, Yurena Gutiérrez, Hèctor Martínez, Laia Mayol, Lucía Medea, Cristina Real, Marina Roman, Carlos Rubio, Io Salmons and Cristina Sospedra Díaz-Maroto. They are simply fantastic, and I thank them every moment we have spent together. I also thank my colleagues in other universities for unforgettable moments: Alejo Alcaraz, Ekaterina Chernova, Wonsuk Jung, Juan Romeu and Melania Sánchez.

Carlos Rubio is one of the luckiest persons in the world because just one person could have been Carlos Rubio, and it happens to be him. I am, however, even luckier because I have spent with him almost every day of the last five years. Carlos has become an indispensable part of my family, and I thank him for everything.

Isabel Castro is known for having been one of the hardest-working PhD students ever. I feel extremely fortunate to know that she is also one of the most extraordinary friends one could have.

Perhaps the most exciting experience in these years as a PhD student was the four months I spent in the University of Cambridge as a visiting student. I thank all the members of the *Rethinking Comparative Syntax* (ReCos) project, specially: Ian Roberts, Theresa Biberauer, Michelle Sheehan, Jenneke van der Wal, Georg Höhn and András Bárány for very fruitful discussions from which this thesis has benefited a lot. It may seem hard to believe, but the most amazing part of my experience in Cambridge was not the city, its beautiful colleges or the wonderful walks in the Backs. The most amazing part was being involved for four months in such a marvellous scientific and personal atmosphere as the ReCoS one.

Cambridge also gave me two gifts I will keep forever: my friendship with Federica Cognola and with Norma Schifano. Federica and Norma are two of the smartest, funniest and kindest persons I have ever met, and I thank them all their love and their care.

***

Finally in these acknowledgments, although he is the very first in all the rest of my life, I thank Jordi Fortuny, to whom this thesis and myself are more indebted than to anyone else. Jordi is not only my favourite linguist, he is also my favourite person.

Contents

List of Figures

LIST OF FIGURES

List of Tables

1

Layout of the dissertation

As the tittle of the dissertation states, in this thesis it will be argued that learning mechanisms active during language acquisition shape the final I-languages acquired. This implies that learning mechanisms as well as learnability constraints delimit variation patterns of natural languages. I will maintain these claims given theoretical considerations and also empirical case studies.

In chapter 2, I will make the proposal explicit, and I will go through some definitions and assumptions that will be relevant for the thesis. I will also make some comments regarding the methodology that has been used.

In chapter 3, I will provide the theoretical arguments for the present proposal. It will be argued that approaching acquisition and linguistic variation as two sides of the same problem has not only desirable theoretical implications, but also allows for a better account of well-known linguistic phenomena. Standard parametric proposals, as well as some of the most famous formal models for parameter setting will be reviewed and it will be demonstrated that they face some relevant shortcomings. Some guidelines about how to develop a plausible theory of acquisition that also makes relevant predictions for variation patterns will be provided.

In chapter 4, a learning constraint derived from simplicity considerations on how the learner acquires affixal morphology is proposed. It will be argued that some synchronic and diachronic linguistic patterns, namely, those concerning the fusional and the agglutinative nature of affixes, derive from

the effects of successive analyses of learners applying this constraint during language acquisition.

In chapter 5, a morphophonological mechanism of data analysis is proposed to be active during acquisition. By means of bootstrapping mechanisms, the values obtained by the learner using the data analyser can be postulated to be triggers for the acquisition of some high-order morphosyntactic properties. By using a methodology like this one, some well-known patterns of natural languages can be analysed in the very same terms as those used by the learner when examining his linguistic input.

In Chapter 6 I point out some open questions and I provide some research lines for further research.

2

Plato's problem and Greenberg's problem

2.1 Aim of the dissertation

The goal of this thesis is to argue, both on theoretical and empirical grounds, that mechanisms in charge of language acquisition define patterns of linguistic variation found across natural languages. This aim amounts to exploring, then, how Greenberg's problem is derivative from Plato's problem.

In general, the literature refers to a "Plato's problem" when some knowledge is not easily explained on the basis of direct learning from the environment, whereby it is necessary to appeal to some other factor, typically innate, to justify it. In this sense, the logical problem of language acquisition, or how it is possible to attain such remarkably complex linguistic knowledge from the environment, is a particular instance of Plato's problem. With the expression "Greenberg's problem" we refer, using Fasanella (2011)'s terminology, to the problem of determining which is the possible range and format of permissible variation among natural languages (although Joseph Greenberg's main objective was not to answer this particular question, this label is used due to Greenberg's effort to discover language universals).

Actually this general research project on how learning mechanisms shape natural languages seems to be similar to what Mendívil-Giró (2012) asks for

in order to go beyond explanatory adequacy (Chomsky 2001b):

> It is quite possible that the human genome specifies very little about what is a possible human language, but it is also true that the genome specifies little about what is a possible human pancreas. In spite of this, each human ends up having a pancreas essentially homogeneous in the species, because the development of this organ (as that of all the others), in addition to being genetically bounded, is strictly constrained by epigenetic and environmental factors and, of course, by principles of development, biochemistry and physical laws. The language organ does not escape this logic (...) However, when biologists talk about the influence of the environment on the developmental process, they are referring principally to the environment of the cell, not to the environment of the organism, which has an undeniable influence, but a far more mediated one. The same applies to language. The environmental (non-genetic) factors that regulate language development are also (but not exclusively) internal to the brain and internal to the language faculty, in the broad sense. *It is precisely in these factors where we should look for language universals (and probably for typological tendencies), and not in the ultimately derived languages, that is, in the different patterns of materialisation* (Mendívil-Giró 2012, 13, *my emphasis*).

According to the quote, and generally according to the biolinguistic enterprise, it is in language-internal areas that we should try to find out why it is the case that linguistic universals and typological tendencies emerge, not in natural languages themselves (or, in other words, in different patterns of materialisation). The main idea pursued in this thesis is that language acquisition is precisely this language-internal domain.

This conception implies that acquisition mechanisms as well as learning constraints not only define a path for learners to acquire language (jointly

with environmental input), but they also shape the possible patterns of language variation and change: in constraining how language is acquired, learning mechanisms define how the possible natural languages finally acquired are.

As suggested before, the main idea of this thesis demands a new approach to language beyond explanatory adequacy. In other words, this general objective expects to answer the question of why language and, specifically, its different patterns of materialisation, natural languages, have their specific format. In the literature one can find studies that give different answers to the question of why language is the way it is: because grammar evolved by natural selection to be efficient in communication (Nowak and Krakauer 1999), because it is optimal to interact with the semantic interface (Chomsky 2001b), or because it shows the structure of general cognitive modules (Lakoff and Johnson 1980), among other hypotheses.

The idea in this thesis is that natural languages have their specific format and shape because language is acquired (in a specific fashion). Actually, given this perspective, we only have to take for granted something that is undeniably true, namely, that language must be acquired. This differs from other hypotheses that have been entertained, such as that language is mainly designed for communication or that language is optimal to interact with semantics, according to which it is necessary to make further assumptions. The research line proposed in this study deserves, thus, a noteworthy role in the biolinguistics agenda.[1]

2.2 Definitions, assumptions and methodology

This project is necessarily more extensive than what can be developed in a single thesis. I will fundamentally be concerned with theoretical arguments in

[1]Following a similar line of research, Culbertson (2010) has argued, using artificial language experiments, that learning biases play a prominent role in the explanation of typological patterns.

favour of this enterprise as well as with empirical case studies that eminently deal with morphosyntactic aspects of natural languages. Ideally this project should also be fruitful in other linguistic domains, though we are not going to explore them in this thesis.

In the next chapters I am going to argue that learning mechanisms shape finally acquired languages, and I am going to explore the effects of two proposed learning mechanisms: a learning constraint, and a mechanism of morphophonological data analysis.

In chapter 3 theoretical arguments will be provided for why it is desirable to conceive a model where Greenberg's problem can be argued to be derivative from Plato's problem. In this theoretical chapter, I will also develop the idea that it is necessary to motivate learning mechanisms on independent grounds before arguing that these mechanisms are responsible for some specific linguistic feature. In other words, it is necessary to argue that some learning mechanism, or constraint, is at work in the acquisitional task and, once this is accomplished, it is possible to argue that this is also the ultimate reason why some variation pattern is found across natural languages. Otherwise, learning mechanisms and constraints would be postulated *ad hoc*, just to derive some concrete linguistic pattern. This is a fundamental aspect of the methodology used throughout this dissertation:

(a) In chapter 4 the proposed learning constraint is derived from simplicity considerations, that is, from computational efficiency reasons, a Third Factor mechanism, which by hypothesis is relevant in the growth of language in the individual (Chomsky 2005).

(b) In chapter 5 I propose a mechanism of morphophonological analysis that operates with two variables relative to morphs, their boundedness and their syntheticity. These two variables, independently whether or not this concrete mechanism is assumed, must be fixed during language acquisition, and therefore, they are motivated on independent grounds.

Actually, what lies beneath both the learning constraint in chapter 4 and the data analyser in chapter 5, as will become clear through this thesis,

is independent from theoretical assumptions and is generally accepted: in order to acquire morphology the learner must isolate minimal meaningful pieces and must give them consistent meanings. As Clark (2001) explains, to acquire morphology, "children must first analyse the structure of words heard in input, identify stems and affixes, map consistent meaning onto them, and then begin to use those stems and affixes in new combinations. This process of analysing form and assigning meaning is a prerequisite in the acquisition of inflectional morphology. It is also a prerequisite in the acquisition of word formation" (Clark 2001, 374).

In more precise terms, then, the learner must a) separate morphs and b) map morphemes onto morphs. Let us provide the definitions of these two concepts, which will be fundamental to this thesis and also are, undoubtedly, central notions in linguistics. However, contrary to what may seem *a priori*, it is not common to find rigorous definitions of these two concepts in the literature. In what follows, we adhere to the definitions provided below:

(1) *Morph*
A linguistic form α, viewed as a string of phonemes, is a morph iff it is meaningful and does not contain any meaningful non-empty proper substring (Fasanella and Fortuny 2012).

(2) *Morpheme*
A semantic grammatical primitive provided by Universal Grammar.

Some clarifications regarding the two concepts defined in (1) and (2) are in order. The definition in (1) captures the idea that a morph is the minimal meaningful unit in natural languages. Given this particular definition, non-meaningful units, such as expletives, are not considered morphs, although their role is, of course, significant for syntax.

The statement in (2) defines a morpheme as the semantic unit provided by Universal Grammar, hence a primitive unit, with grammatical content. According to this definition, given that lexical meanings that are expressed, for instance, by lexemes *dog* or *pen* are not grammatical but lexical, *dog* and *pen* are not considered morphemes. This particular definition, which

is not standard (notice that it is usual to conceive of both "grammatical" and "lexical" morphemes), allows us to precisely state the role of Universal Grammar in dealing with a concrete aspect of meaning: Universal Grammar provides semantic grammatical primitives, which we call morphemes. As a consequence, we assume that Universal Grammar does not provide the learner with any concrete lexical meaning.[2]

Given these definitions, a morph can be associated with at least one morpheme or with at least one lexical meaning, whereas a morpheme can be instantiated by a particular morph, along with other morphemes or not, or can be covert.

Some other assumptions that are relevant here, and in the remaining chapters, are those concerning Universal Grammar, linguistic variation and the poverty of stimulus argument.

In this thesis, Universal Grammar (UG) is conceived as the human device that enables language acquisition, leaving aside to which extent its content is exclusively human and/or specifically linguistic (Hauser et al. 2002).

Regarding linguistic variation, we draw on Roberts (2007)'s observation that there are parameters of a different nature. When explaining a parameter dealing with negative concord, and after going through some other parameters concerning verb-movement and dropped subjects, Roberts remarks: "parameter D [the negative concord parameter] is rather different from the others discussed here, as it does not directly concern word order, but rather variation in the feature-content of certain classes of lexical items. It is also more closely connected with semantics than the other parameters we have looked at" (Roberts 2007, 81).

Although I will not commit myself to the classical notion of "parameter" (see chapter 3), I share Roberts' idea about the (at least apparent) two observable types of morphosyntactic variation in natural languages: first,

[2]With regard to this point, it is worth considering whether all possible morphemes are expressed in all languages. We will assume that they are not. Following Chomsky (2001), we take Universal Grammar to contain a universal pool of features, and particular languages to express a subset of them. Nonetheless, our proposal is not contingent on this particular assumption.

what may be called "feature-content" variation and, second, the so-called "word-order" variation. We can say that the first type of variation concerns how morphs and morphemes are assembled and what morphosyntactic consequences it produces, whereas the second type of variation has to do with the possibilities of the relative position of categories, their possibilities of being omitted and the configuration where they are licensed.

In this thesis I will only be concerned with the feature-content type of variation, and I will argue that the way in which learning mechanisms guide the process of identifying morphs and assigning morphemes to them shapes the final acquired I-languages in some predictable morphosyntactic terms. I will also argue that this process not only has consequences for synchronic patterns of morphosyntactic variation, but also for diachronic ones, that is, for patterns of language change.

It is worth noting at this point that the Uniformity Hypothesis, proposed by Chomsky (2001), is assumed:

(3) *Uniformity Hypothesis (UH)*
In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances (Chomsky 2001, 2).

Focusing on the type of variation I am dealing with in this thesis, I will propose that a subset of those "detectable properties of utterances" is relative precisely as to how morphs are organised and how morphemes map onto morphs. Several morphosyntactic patterns of variation derive from this morphological assembling.[3]

I adhere to the Poverty of Stimulus argument inasmuch as I conceive language acquisition as an instance of Plato's problem. At the same time I take language to be a biological entity that is subject to both domain-general factors and natural laws in its phylogenetic and ontogenetic development.

[3]As has already become clear, when referring to linguistic variation I consider just morphosyntactic variation, not prosodic or phonological variation, which are of course other types of linguistic variation and also "detectable properties of utterances" that I am not going to discuss.

Given that, the general macro-purpose is the usual one in generative linguistics, namely, trying to reduce the *a priori* postulated language specific constructs in favour of discovering more fundamental underlying components when accounting for linguistic phenomena (Berwick et al. 2013).

Throughout this thesis the labels 'child', 'learner' and 'Language Acquisition Device (LAD)' will be used indistinguishably, and will correspond to the idealised learner, endowed with UG, whose task is acquiring a natural language.

3

Towards a discovery procedure for language acquisition

Since the beginning of the Generative Grammar enterprise, an apparently insurmountable tension emerged between what has been called *descriptive adequacy* and *explanatory adequacy* (Chomsky 1965): whereas cross-linguistic variation is a striking phenomenon and languages diverge from one another to a remarkable degree, the Faculty of Language must be fundamentally homogeneous given the uniform and stable nature of the process of language acquisition. A proposal meets descriptive adequacy when it describes and predicts linguistic facts known by natives speakers, and it meets the deeper condition of explanatory adequacy when it accounts for the acquisition of such a knowledge.

In the mid-20th, once the logical problem of language acquisition was brought into consideration (Chomsky 1959), Generative Grammar tried to combine the analysis of syntactic variation with an eminently uniform conception of the Faculty of Language. One of the most important contributions of this research line is the new understanding of UG, the human device that enables language acquisition. In particular, following the conception suggested by the Principles and Parameters (PP) framework (Chomsky 1981), UG contains a finite set of principles each of which is associated with one parametric variation, or parameter, that has to be fixed through experience.

3. TOWARDS A DISCOVERY PROCEDURE FOR LANGUAGE ACQUISITION

By means of *principles*, UG becomes an apparently enough constrained device capable of permitting language learning, meeting explanatory adequacy, while linguistic variation could be accounted for by appealing to the different fixation of *parameters*, meeting descriptive adequacy.

Therefore, the conception of a parametrised UG or, more concretely, the very idea of parameters, was proposed to address the apparent mismatch between the answers to both Plato's problem (or how are natural languages learned?), and Greenberg's problem (or what is the degree and format of permissible variation?).

As Yang's statement summarises "the theory of parameters is charged with two ambitious missions –to provide a theory of the languages of the world and the language of the child– in a single stroke" (Yang 2006, 131).

Although this was the original conception of parameters, concrete parametric studies have mainly been focused on just one of the two sides of the problem, as will become clear in the next sections of this chapter.

On the one hand, traditional parametric proposals as well as some of the more recent minimalist approaches (Chomsky 1993, 1995) and subsequent work, understanding Minimalism as a second stage of the PP framework) have stressed the role of parameters in describing cross-linguistic morphosyntactic variation, but it is crucial to keep in mind that a model of PP must also provide the elements that guide the learner in the process of interpreting the data he receives in terms of linguistic evidence in order to attain a particular I-language.

On the other hand, some other approaches, which can be called for the sake of clarity 'learnability approaches', have focused on how a theory of parameters can be accommodated to concrete language learning algorithms. However, these learning algorithms operate with extremely idealised artificial languages, and as a result these proposals are not able to handle some important facts about natural languages.

The rationale behind the argumentation that will be developed in this chapter is that there are major advantages in conceiving a theory where these two questions, acquisition and variation, are approached as two sides of the same problem. This view is not only desirable on theoretical and

empirical grounds, as will be demonstrated throughout this thesis, but it is also a way of providing the theory with self-contained elements of evaluation. In other words, if morphosyntactic variation is considered to be constrained by learning mechanisms, it is possible to evaluate proposed mechanisms of language acquisition regarding their predictions about what is the degree and format of variation among natural languages. Likewise, proposals concerned with establishing patterns of linguistic variation can be assessed as to the assumptions they made regarding how this variation is acquired.[1]

The structure of this chapter is the following: in section 3.1, I show how standard parametric proposals face some problems regarding language acquisition by neglecting three proposed learnability conditions; this section is based on Fasanella and Fortuny (2012). In section 3.2, I argue that, in general, learnability approaches are not able to derive some well-known facts about the nature of linguistic variation. In section 3.3, I provide some relevant connections between the main insights of the previous two sections; in section 3.4, I introduce some guidelines about how to put forward an approach to language acquisition that follows learnability considerations and that is capable of deriving important facts about patterns of linguisitic variation; or, in other words, how would a theory answer both Plato's problem and Greenberg's problem. Section 3.5 is the conclusion of this chapter.

3.1 Focusing on Greenberg: a critical review of some parametric approaches

3.1.1 Macroparametric proposals

The type of parameters mainly proposed within the *Government and Binding* approach (Chomsky 1981), where each parameter was related to a principle, but also in some recent studies, is known as *macroparameter*, since Baker

[1]Pinker (1979) notes that Hamburger et al. (Hamburger and Wexler 1975, Wexler and Hamburger 1975, Culicover and Wexler 1977) have clarified and justified this central tenet in transformational linguistics, namely, that considerations of language learnability can dictate a choice between rival linguistic theories.

(2008). Macroparameters are conceived as Boolean variables. For the study of language variation, this implies that all languages can be examined with respect to their particular parametric value in relation to a principle.[2] Moreover, it is proposed that if a language has fixed a parameter with a given value, then it shows some syntactic properties related to that specific parameter, whereas if a language fixes the parameter with the opposite value, then it will not show these properties. Regarding acquisition, it is assumed that the child fixes each macroparameter by inspecting its language environment. Likewise, if the child fixes the parameter with a given value, he is able to acquire those associated properties, known as the *clustering properties*, which usually are conceived as more abstract and less accessible than parameters themselves.

A classical example of parameter postulated in the eighties was the Pro-Drop Parameter or Null Subject Parameter (NSP), which was related to the principle that specified that all sentences need a formal subject, argumental or not. However, given that in some languages this subject can be superficially null, the NSP was proposed. When the NSP is fixed with a positive value in a given language, that is, when it is possible to omit the subjects, like in Italian or Catalan, then this language displays the following clustering properties (Chomsky 1981, 240):

(1) Clustering properties of the NSP:

(a) Missing subject.

(b) Free inversion in simple sentences.

(c) Long *wh*-movement of subject.

(d) Empty resumptive pronouns in embedded clauses.

(e) Apparent violations of the *[that-t] filter.

[2]It is assumed by default that all languages have a positive or negative value for all parameters, although some works propose the opposite (Newmeyer 2005).

These are the illustrations of properties in (1) in Italian, which, unlike English, is a null subject language (Chomsky 1981, 240):

(2) Italian examples of the clustering properties in (1):

(a) Ho trovato il libro (I found the book).

(b) Ha mangiato Giovanni (Giovanni ate).

(c) L'uomo [che mi domando [chi abbia visto]] (with the interpretation "the man x such that I wonder who x saw").

(d) Ecco la ragazza [che mi domando [chi crede [che possa VP]]] (This is the girl who I wonder who thinks that she may VP).

(e) Chi credi [che partirà] (Who do you think -that- will leave).

The relation between the positive fixation of a parameter and its related clustering properties was perceived as an important step forward in solving Greenberg's problem, since it allows the theory to relate *a priori* unrelated syntactic patterns (as those in (1)) with only one property of a given language, namely, its specific parametric value with respect to one principle.[3]

Regarding the process of language learning, the fixation of a parameter by the child supposedly has *cascade effects*: once the parameter is set, the clustering properties are assumed to be easily and almost directly acquired.

Apart from the NSP, which is sometimes considered as a medioparameter,[4] perhaps two of the most influential macroparameters proposed in the literature are the Polysynthesis Parameter and the Compounding Parameter.

[3]See for example Gilligan (1987) and Nicolis (2008) for a critical review of the prevalence of the clustering properties related to the positive setting of the NSP.

[4]Baker (2008) proposes this label, as well as the famous "macroparameter" and "microparameter", in order to differentiate parameters regarding three aspects: their supposed locus in grammar, their associated clustering effects and the range of languages that, once compared, led to their discovery. Medioparameters would be between macro- and microparameters in the sense that they are syntactic but sometimes do not affect the core of grammar, they have clustering properties but not as general in their effects as macroparameters and, finally, they were discovered neither comparing dialects nor different language families, but just relatively related languages.

3. TOWARDS A DISCOVERY PROCEDURE FOR LANGUAGE ACQUISITION

The Polysynthesis Parameter was proposed by Baker (Baker 1996, 14) and states the following:

(3) *The Polysynthesis Parameter*

> Every argument of a head element must be related to a morpheme in the word containing that head.[5]
> Yes: Mohawk, Nahuatl, Mayali, etc.
> No: English, French, Chichewa, etc.

The argued properties associated with the positive fixation of the Polysynthesis Parameter that consequently languages like Mohawk or Nahuatl should show are listed below (we will come back to the Polysynthesis Parameter in the next sections and chapters, specially in chapter 5):

(4) Clustering properties of the Polysynthesis Parameter:

(a) There is subject and object agreement on the inflected verb (person, number and gender).

(b) There is possessor agreement on the noun.

(c) Inflections are obligatory and fixed in position.

(d) Word order is free, noun phrases can be omitted and discontinuous constituents are possible.

(e) There is noun incorporation.

We can observe some of these properties in this example given by Baker (Baker 1996, 23) from Southern Tiwa, a Tanoan language:

(5) Bi-seuan-m$\tilde{u}$-ban
1sgS.3plO-man-see-PAST
'I saw the man.'
(Allen et al. 1984, 295)

[5]We mantain the original formulation by Baker (1996), but note that what Baker considers *morphemes* are, in our terminology, *morphs*.

The subject is solely expressed by a morph in the verb that at the same time agrees with the object of the sentence, *bi* (1sS.3pO), and the noun *seuan* ('man') is incorporated in the verb *mũ* ('see').

The Compounding Parameter was proposed and defined by Snyder as follows (Snyder 2001, 328):

(6) *The Compounding Parameter*

> The grammar disallows* or allows formation of endocentric compounds during the syntactic derivation [*unmarked value].
> Allowance: English, Estonian, Mandarin, etc.
> Disallowance: Spanish, Serbo-Croatian, Javanese, etc.

The clustering properties that are associated with the Compounding Parameter in (6) are the complex-predicate constructions in (7), which are possible in languages that allow the formation of endocentric compounds:

(7) Complex-predicate constructions associated to a concrete fixation of the Compounding Parameter (Snyder 2001, 325):

(a) Resultative: John painted the house red.

(b) Verb-particle: Mary picked the book up/picked up the book.

(c) *Make*-causative: Fred made Jeff leave.

(d) *Put*-locative: Bob put the book on the table.

(e) *To*-dative: Alice sent the letter to Sue.

(f) Double-object dative: Alice sent Sue the letter.

Therefore, according to Snyder's proposal, if a language has truly novel (non-lexical) Noun-Noun compounds without overt morphological connectors, it should show the complex-predicate constructions in (7), whereas a

language that does not allow the formation of such compounds should not show these constructions.[6]

The usual criticism against macroparameters, commonly found in the minimalist literature, appeals to the argument of evolutionary plausibility. It is claimed that the more UG is overspecified with articulate and intricate networks of parameters, the less plausible is an account of its supposedly recent evolutionary origin (Chomsky 2005, Roberts and Holmberg 2010). Another criticism that does not resort to evolutionary arguments, but deals with learnability considerations, is in order: although macroparameters may seem elegant in systematising cross-linguistic variation (but see the discussion in section 3.2 on mixed patterns), they are not adequate to define formal models for language acquisition, since they exhibit what may be called the Locality Problem:

(8) *Locality Problem*
In order to fix the value of a macroparameter the learner should analyse the data he receives in a global and transverse way, since macroparameters are defined on highly general properties that are spread across the target language.

I illustrate this problem with the Polysynthesis Parameter. For the learner to fix the value of such a general statement like (3), he should be able to determine what the arguments and the heads of a given sentence are, what arguments depend on what heads, by means of which morphs arguments are realised in the word containing the heads, among many other properties. Moreover, all these analyses should be carried out for every sentence or piece of input the learner would receive and always before the fixation of the parametric value. Nevertheless it is obvious that the child cannot have access to a sufficient number of tokens of all relevant types of constructions

[6]Snyder himself recognises that some languages that do not allow the formation of endocentric compounds show, however, some of the clustering properties in (7). Actually he considers the appearance of noise in data an unavoidable fact when dealing with a macroparameter (Snyder 2012). See Son (2007) for a critical review of the clustering properties associated with the Compounding Parameter and its cross-linguisitic validity.

of its language environment at once, as we linguists have when examining the properties of a given language. This does not mean that the child does not receive enough data to set parameters, but rather that, in order to set the value of a global macroparameter such as the Polysynthesis Parameter, it is necessary to determine properties of a multitude of less general parameters tacitly compressed in the macroparameter; for example, whether the nominative morpheme is expressed in the word containing the verb, whether the accusative morpheme is expressed in the word containing the verb, and so on. In other words, it is hardly conceivable that the LAD can set the value of a macroparameter unless less general properties are set before.

This entails that the the global procedure required given a macroparametric acquisition approach is highly unlikely and, contrary to what has been claimed, macroparametric schemata, as the one in Baker (2001), do not constitute plausible learning paths, since they do not adhere to the following efficiency learnability condition:

(9) *Atomicity Condition*
Parameters must be atomic, they cannot be clusters of properties.[7]

This criticism also applies to most considered medioparameters, and in general to how standard parametric models have been developed so far, given that the definition of parameters has been guided by the ideal of compressing multiple properties in a single and more abstract property. Inasmuch as the search for parameters is conceived in this way, the Atomicity Condition is automatically neglected, thereby leaving learnability considerations aside.

3.1.2 Microparametric proposals

The advent and development of Minimalism (Chomsky 1993, 1995) involve significant methodological and conceptual consequences regarding the treatment of morphosyntactic variation. Mostly motivated by Richard Kayne's

[7]This learnability condition is reminiscent of Kayne's general conjecture "every parameter is a microparameter" (Kayne 2005b, 10), but as we will argue this condition is not sufficient.

work, multiple studies focused their attention on closely related linguistic variants. As argued, by comparing closely related languages instead of languages that belong to different families, it is possible to observe in detail the specific point of variation that one wants to pay attention to without so much noise (Kayne 2000). This approach to linguistic variation leads to the concept of *microparameter* (Baker 2008).

Since microparameters specify very concrete and specific points of variation, it is assumed that all of them could not be codified in UG, since it would become an excessively intricate and totally implausible device. Instead, what is usually claimed from this perspective is that all morphosyntactic variation, encoded in microparameters, could be reduced to general schemata, perhaps to only one, which are concerned with formal features. As a result, only the schema(ta) would belong to UG and each particular case of variation would have to be accomodated to it. In more recent studies, though, it has been proposed that parameter hierarchies are not primitives of UG and arise, at least partially, from Third Factor pressures (Biberauer et al. 2013).

Emphasising the twofold goal of parametric theory, microparametric schemata are understood both as potential learning paths and as typological biases. One of the first attempts to develop an explanation along these lines is the work initiated by Gianollo et al. (2008). These authors claim that a large number of parameters related to the nominal phrase can be reduced to five independent parameter schemata (Gianollo et al. 2008, 119):

(10) Parametric schemata for variation in the nominal domain:

(a) Is F, F a functional feature, grammaticalised?

(b) Is F, F a grammaticalised feature, checked by X, X a category?

(c) Is F, F a grammaticalised feature, spread on Y, Y a category?

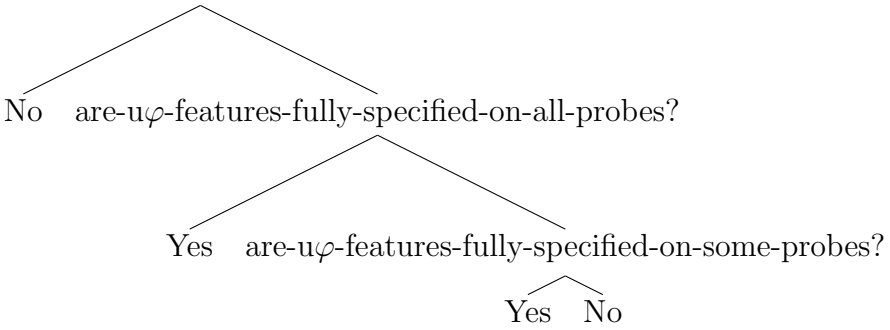
(d) Is F, F a grammaticalised feature checked by X, strong (i. e., does it overtly attract X)?

(e) Is F, F a grammaticalised feature, checked by a category X^o?

The schemata in (10) is supposed to be part of UG, along with principles. Therefore it becomes unnecessary to assume that all microparameters that can be accommodated to one of the statements in (10) are specified in the initial state of language.

Another example of parameter schema is the one proposed by Roberts and Holmberg regarding null arguments. They argue that the typological generalisations concerning this phenomenon can be accommodated to a specific schema (Roberts and Holmberg 2010, 49), simplified as follows:

(11) are-uφ-features-obligatory-on-all-probes?



If the first question is answered by the child with 'no', then he will end up having a radical pro-drop language; if the second is a 'yes', then his language will be polysynthetic; regarding the last question, a 'yes' will describe a non-pro-drop language and a negative answer will point to the different types of null subject languages.

A simpler schema is proposed by Boeckx with the aim of summing up some of the most relevant parameter schemata in the literature (Boeckx 2011, 215):

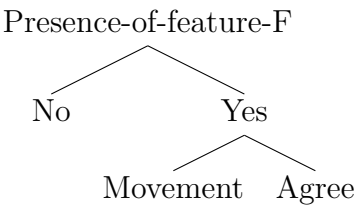
(12) General parameter schemata by Boeckx (2011):

(a) Is F present/active in the language? Yes/No

(b) If Yes, does F give rise to Movement, or simply Agree(/Merge)?

The proposal in (12) could also be expressed as a hierarchy of feature behaviour since it presents a nested structure:

(13)



A commonly noted advantage of parametric schemata is that they do not imply the idea of particular languages as a block, like macroparameters do when postulating that Catalan is a null subject language but English is not, for example. Schemata apply to a specific feature, not to a whole language. With this method it is possible to explore linguistic variation within languages with the same type of tools used to study linguistic variation across languages, since it is assumed that parametric variation in the two cases are not of a different type, since the involved elements are logically the same (Kayne 2005a).

Another conceptual assumption that has become notably relevant in the treatment of linguistic variation within Minimalism is the Lexical Parameterisation Hypothesis, devised by authors like Borer, Manzini, Wexler and Webelhuth in the eigthties and the early nineties:

(14) Values of a parameter are associated not with particular grammars, but with particular lexical items (Newmeyer 2005, 54).

Actually this is the current common minimalist view concerning morphosyntactic variation. In particular, most of the proposals assume the hypothesis called by Baker the Borer-Chomsky Conjecture (BCC), which states:

(15) All parameters of variation are attributable to differences in the features of particular items (e.g., the functional heads) in the lexicon (Baker 2008, 353).

As a result, most minimalist approaches deal with linguistic variation trying to find the relevant parametric schema which has to be part of UG, and trying to derive morphosyntactic variation between languages from points of featural divergence in functional heads.

3.1 Focusing on Greenberg: a critical review of some parametric approaches

As pointed out in the literature, Third Factor mechanisms should play an important role in the general working of microparametric schemata. Third Factor mechanisms are part of the mechanisms that Chomsky argues that enter into the growth of language in the individual, which are the following (Chomsky 2005, 6):

(a) First Factor: the genetic endowment for language in human beings.

(b) Second Factor: experience.

(c) Third Factor: principles not specific to the Language Faculty, that include

- principles of data analysis.
- principles of structural architecture and developmental constraints, including principles of efficient computation.

Principles of general conservativity, considered Third Factor mechanisms, are assumed to strongly guide the learner in the process of going through the learning path that microparametric schemata define (we will turn to this point in section 3.4). This intuition is expressed by Holmberg and Roberts (2010) in the following terms:

> Acquirers, and therefore languages, favor a high position in the network. For general reasons of computational conservativity, the acquirer is always trying to stop, as it were. Acquirers only move down the tree if forced to by the PLD.

Nonetheless, it is crucial to notice that microparametric schemata rely on excessively abstract syntactic notions, like the checking operation in (10), the probe-goal relation in (11) or the Agree operation in (12), which the learner cannot be assumed to directly detect on the basis of linguistic input. Consequently, it can be argued that acquirers would not even be able to begin with the schema if guided by PLD. These approaches should clarify, for instance, how the infant could determine when a grammatical feature is

checked by a category, what counts as a probe, or when an Agree relation takes place, by inspecting the PLD. Note that, even if we assume that all these operations and formal entities are part of UG, it is necessary to provide the means of connecting these prior notions to the analysis of linguistic evidence. In sum, microparametric schemata face the so-called Linking Problem, which has been already observed in the acquisition literature:[8]

(16) *Linking Problem*
Parameters are defined over abstract linguistic entities such as verbs, nouns and pronouns, so the infant still faces the problem of linking these abstract mental representations to actual physical entities in the speech signal (Gervain and Mehler (2010) 194, cfr. Pinker (1984)).

Accordingly, and contrary to what is commonly claimed, microparameters are not plausible learning paths useful for the child in order to acquire language. More precisely, although microparametric schemata satisfy the Atomicity Condition in (9), they fail to satisfy the following learnability condition:

(17) *Accessibility Condition*
Parameters must be set by directly inspecting phonological and morphological properties of utterances.

In order to solve the Linking Problem, the existence of *cues* has been postulated. Cues are understood as accessible parts of the linguistic input that are associated with parameters and that enable their valuation by the learner. Models assuming cues (Dresher and Kaye (1990), for example) defend the idea that UG provides the learner with all the parameters and with all the cues associated one by one to them. Although this is an attempt to solve the problem of linking the input environment with the abstract entities parameters refer to, this move is a complication of what UG must consist of (principles, parameters and also cues), that should be avoided if possible

[8]The same problem has also been labeled the Epistemological Problem (Dresher 1999) and the String-to-structure problem (Sakas and Fodor 2001).

(see also next section 3.2 for more details about Dresher and Kaye (1990)'s model).

A third general learnability problem shared by some macro- and micro-parametric proposals is that one of the two values of a parameter (or of a parametric statement) would be fixed on the basis of negative evidence. We call this problem the Indetermination Problem:

(18) *Indetermination Problem*
One of the two values of certain parameters cannot be determined on the basis of positive evidence.

Consider for concreteness the first statement of the microparametric schema in (12): "Is a feature F present/active in the language? Yes/No" . Let us reflect about how the child could proceed to value this particular statement. We could be tempted to believe that the child could set its positive value if he were provided with the appropriate positive evidence, say a relevant number of sentences displaying an overt mark of feature F. However, it is unclear how he could discover that F is not phonologically realised in the target language. In order to show that this parametric statement is part of a plausible learning path it is necessary to propose a mechanism to compensate for the lack of negative evidence. For instance, one could try to find out how many sentences the learner should receive before concluding, with no direct positive evidence, that F is not phonologically realised. This is known as the use of 'indirect negative evidence' (Berwick 1985). The same problem extends to other microparametric statements as well as to some macroparametric proposals. Another example could be one of the five microparameters in (10): "Is F, F a grammaticalised feature, checked by X, X a category?" If the learner can positively detect that X checks F, then he can assign a positive value to this microparameter, but one must ensure that the learner will not search for positive evidence for ever when the target language does not display this checking operation. Seemingly, if a leaner were trying to assign a value to the so-called Polysynthesis Parameter (3), he would search endlessly for positive evidence if the target language does not show that every argument of a head is related to a morph in the word containing that head.

3. TOWARDS A DISCOVERY PROCEDURE FOR LANGUAGE ACQUISITION

It is possible to avoid the Indetermination Problem in (18) by assuming that the value of a parameter that cannot be set on the basis of positive evidence is the default value. The parameter would initially be set to this default value and would only be switched to the non-default value on the basis of positive evidence. Actually, the first 'parameter' proposed in the literature (Rizzi 1982, footnote 25), the possibility of having different bounding nodes for subjacency, is conceived in this fashion because of learnability considerations: the more restrictive option (NP and S are bounding nodes) is the default value and is only partially abandoned by the learner if data shows the contrary (S is not a bounding node in Italian, for example).[9]

In brief, if parameters are set through experience, formulating parameters that cannot be fully determined on the basis of positive evidence introduces complications from a learnability point of view that must be avoided. It is also conceptually implausible that UG is constituted of useless instructions for language acquisition, which cannot be easily determined on the basis of linguistic input. This rationale is expressed in terms of the following condition:[10]

(19) *Positive Evidence Condition*
Both values of a parameter must be set on the basis of positive evidence.

The three conditions formulated in this section, the Atomicity Condition (9), the Accessibility Condition (17), and the Positive Evidence Condition (19) are directly deduced from learnability considerations that have traditionally

[9]This move, jointly with some parameter ordering, can also be a way of avoiding local maxima, that is, the situation where the learner is stuck in an absorbing state given a parameter space, as in Gibson and Wexler (1994)'s model, that will be briefly reviewed in the next section.

[10]In a very general sense, this Positive Evidence Condition is covered by the Accessibility Condition, which states that parameters must be set by *directly* inspecting phonological and morphological properties of utterances. Nonetheless it is worth making this condition explicit because a great number of parameters found in the literature have a value that cannot be set on the basis of positive evidence in the PLD, leaving aside whether or not they appeal to abstract and non-detectable entities.

been overlooked. Indeed these three conditions are no more than principles governing data analysis used in language acquisition, and thus would be part of Third Factor mechanisms (Chomsky 2005).

3.2 Focusing on Plato: a critical review of some learnability approaches

In this section, I will review four representative proposals of what can be called 'learnability approaches', that is, studies mainly focused on the formulation of learning algorithms for language acquisition by means of parameter setting: Gibson and Wexler (1994), Niyogi and Berwick (1996), Dresher (1999) and Yang (2010). The goal of this review is to sketch very briefly the main features of each model in order to evaluate them regarding Greenberg's problem. I will argue that they cannot be taken to be models for the acquisition of natural languages because the artificial languages they operate with are excessively idealised to handle very spread patterns of variation. Particularly, we will pay attention to three specific properties of morphosyntactic variation. Some of these models also neglect some learnability conditions proposed in the last section, like the Accessibility Condition in (17), but we will focus on some consequences regarding linguistic variation patterns.

Gibson and Wexler (1994) explore how a concrete algorithm of parameter setting is related to the triggering data. As they argue, it is tacitly or overtly assumed in parametric studies that something in the input will be used as a trigger for any concrete parameter for the learner to be able to value it. Triggers are defined as follows:[11]

(20) *Trigger.* Given values for all parameters but one, parameter P_1, a local trigger for value v of parameter P_1, $P_1(v)$, is a sentence S from the target grammar L such that S is grammatical if and only if the value for P_1 is v.

[11]Indeed they defined two types of triggers, global and local, but we will concentrate on the local type for the sake of simplicity since this distinction does not affect the general argumentation.

Therefore a trigger for a given parameter is the sentence that points to the correct setting of that parameter and, thus, forces the learner to value it correctly.

As the authors explain, given the definition of trigger in (20), there can be no possible triggers for subset values of concrete parameters. A value_1 of a parameter is in a subset relation with the other value_2 of the parameter when sentences that can successfully be analysed with the value_1 can also be analysed with the value_2, but not viceversa. We can exemplify this case with the NSP presented in the previous section. Catalan is a null subject language and English is not. As we see in the Catalan examples, the subject may be omitted (21) or may be not (22), whereas in English it cannot:[12]

(21) Canta
sing.3sg
'*(He/She) sings.'

(22) En Joan canta
Det. Joan sing.3sg
'John sings.'

Given this situation, sentences like (21) will be triggers for the positive valuation of the NSP, whereas sentences like 'En Joan canta' or 'John signs' can successfully be analysed with both values of the NSP. Accordingly, the English learner will not have a real trigger in his input to fix the NSP negatively. All things being equal, only nonsubset values of parameters have triggers. It is for this reason that Gibson and Wexler restricted themselves to parameters that do not contain subset values (we will come back to the subset/superset situation at the end of this section).

In this work, Gibson and Wexler also formalise a Triggering Learning Algorithm, that is, an error-driven algorithm that the learner uses to relate triggers in the input sentences with the fixation of parametric values to acquire language:

[12]The general picture is even more complicated if we take into consideration sentences where the subject pronoun is realised, as in *ELL canta* 'He sings', which are emphatic or contrastive.

(23) *Triggering Learning Algorithm (TLA).* Given an initial set of values for n binary parameters, the learner attempts to syntactically analyse an incoming sentence S. If S can be successfully analysed, then the learner's hypothesis regarding the target grammar is left unchanged. If, however, the learner cannot analyse S, then the learner uniformly selects a parameter P (which probability $1/n$ for each parameter), changes the value associated with P, and tries to reprocess S using the new parameter value. If analysis is now possible, then the parameter value change is adopted. Otherwise, the original parameter value is retained.

The TLA is conservative in the sense that it changes its actual fixation step by step, that is, parameter by parameter, and only if the change improves the analysis of the relevant sentence.

Once these important notions are formalised, Gibson and Wexler explore how the TLA would work given a very simple parameter space, with just three parameters: two regarding word order (one for the order of the complement, one for the order of the specifier) and another regarding verb second phenomenon.

Their main finding is that notwithstanding the small number of parameters of their sample, the learner will not always converge on all possible target grammars, due to two reasons. First, the authors show that there will be no triggers for all possible target grammars, understanding target grammar as those eight possible grammars generated by the different combination of binary values of the three parameters considered. This means that for some target grammars, there will be no trigger: an input that is grammatical if and only if a parameter has a certain value given this concrete parameter space. Second, they prove that using the TLA the learner will get stuck in *local maxima*, that is, in absorbing states different from the target grammar (which should always be an absorbing state) from which he cannot escape given the linguistic input available.

At this point, Gibson and Wexler argue that, instead of abandoning the idea of linguistic triggers and parameters, two changes should be made to the

theory in order to avoid the problems mentioned above. Although they consider other possible solutions as well, they seem to prioritise the following: a) to postulate a default initial grammar, and b) a parameter-ordering sequence, such that the learner never gets stuck in local maxima states. Accordingly, if the parameter space is not free but there is an ordering-sequence with default values for parameters, the learner can override the lack of triggers and the possibility of getting stuck in an absorbing state, and he will be able to set all parameter values.

Niyogi and Berwick (1996)'s study is focused on the question of the convergence time that will be required to reach the target grammar given a learning algorithm; or, in other words, they give special attention to the question of how many positive examples it will take to set all parameters. These authors take as a starting point the TLA proposed by Gibson and Wexler (1994), and the three parameters scenario Gibson and Wexler describe, and Niyogi and Berwick mathematically formalise the TLA algorithm as a Markov chain. Using this formalisation Niyogi and Berwick can precisely measure how many examples will lead to the convergence on the target grammar. The result they obtained is between 100 and 200 positive sentences, a psychologically plausible number.

As for the question of the subset/superset values of parameters, they follow Gibson and Wexler (1994) and only consider a parameter space where these types of parameters does not arise, that is, where it is not the case that a parametric value is a subset of another one.

Due to the stochastic method Niyogi and Berwick apply, they can obtain a surprising result: they demonstrate that given a learning algorithm as the TLA, the learner will not necessarily converge always on the target grammar, even if it is possible to avoid the local maxima states somehow, for instance introducing the changes pointed out by Gibson and Wexler. As Niyogi and Berwick express it: "the existence of a chain of 'triggers' from a source to target language (grammar) does not suffice to guarantee learnability" (Niyogi and Berwick 1996, 176).

In order to solve this problem, the authors propose some changes regarding the original TLA. Instead of having the constraint of changing just

one parametric value at the same time, as in the original TLA, Niyogi and Berwick give the learner the possibility of changing an indefinite number of parametric values when he cannot analyse a concrete sentence. For instance, imagine a parameter space with two parameters where parameter_a is fixed with the value_a^1 and parameter_b is fixed with the value_b^1. Given the original dynamics of the TLA, the learner has the possibility to change just one of these two parametric values at once. For example, he could change the value of parameter_a from value_a^1 to value_a^2 and then he has to stop and has to analyse its input with this concrete parameter setting. However, assuming Niyogi and Berwick (1996)'s modifications, the learner could change the values of parameter_a and parameter_b at the same time before analysing the input: from value_a^1 to value_a^2, and from value_b^1 to value_b^2.

This move can eliminate the local maxima insofar as the learner is able to conjecture hypotheses far from its current one in the parameter space. As a consequence, the parameter space has only one absorbing state, the target grammar. Apart from that, Niyogi and Berwick (1996) prove that a random algorithm (start at a random point and if the sentence cannot be analysed, move to another random point) is actually faster than the TLA in terms of the convergence time needed to reach the target language.

Another appealing aspect of Niyogi and Berwick (1996)'s proposal is that they use not only artificial constructed input, but they also examine the utility of their Markov model using real language distributions extracted from the CHILDES database, where they obtain roughly the same results, about 100 simple sentences are needed to reach the valuation of parameters.[13]

Dresher (1999) investigates how a cue-based parametric model (further developed in Dresher and Kaye (1990)) can solve two important learnability problems. The cue-based model was proposed mainly to address the acquisition of phonology, but Dresher claims that it can also be applied to the acquisition of other areas of linguistic knowledge. The principal properties of this model are the following (slightly simplified):

[13] As the authors explain, these results concern parameters of word order and not the verb second parameter, whose valuation seems more difficult to be guaranteed given their system.

(a) UG associates every parameter with a cue, which is not an input sentence or form but something that can be derived from input.

(b) Cues are appropriate to parameters in the sense that they reflect a fundamental property of the parameter.

(c) Parameter setting proceeds in a (partial) order set by UG: this ordering reflects dependencies among cues and specifies a learning path.

(d) A parameter that has a default state remains in it until the learner detects its cue, which acts as the trigger to move to the marked setting. Symmetrical parameters may have positive cues for both values.

(e) Cues become increasingly abstract and grammar-internal the further along the learning path they are.

Therefore, according to the cue-based learning model, language acquisition proceeds through an ordered path that the learner follows given specific cues related to any single parameter that become progressively more abstract.

As we have indicated above, Dresher (1999) emphasises that the cue-based learning model can overcome two important problems regarding language acquisition that other learning models cannot successfully solve. One is the Epistemological Problem, that can be roughly equated with the Linking Problem stated in section 3.1 of this chapter. The other problem is what Dresher calls the Credit Problem: "When there is a mismatch between a target form and a learner's grammar, there is no way of reliably knowing which parameters/constraints must be reset to yield a correct output" (Dresher 1999, 28). In other words, the Credit Problem states that when a learner hypothesises a grammar that does not succeed in analysing the target input, he will not have information about the nature of the error. Leaving aside the relevance of the second problem in computational models of language acquisition (recall that Niyogi and Berwick (1996) demonstrate that a random algorithm in a given parametric space is able to converge), Dresher's observation regarding the Credit problem may be of interest when facing real acquisition scenarios.

In the context of the cue-based model (and not in other learning models, as Dresher argues in this study), these two problems can be resolved. The Epistemological problem is solved by means of the learning path itself, that becomes increasingly abstract in the course of acquisition. The learner is supposed to begin with the valuation of parameters regarding prosodics and phonology before the valuation of syntactic parameters starts. The Credit Problem does not arise because the learner always knows what to look for to set a parameter due to the fact that each parameter is associated with a cue.

In this study, Dresher defines two types of possible learners, the batch learner and the incremental learner. The batch learner is the one that firstly collects all necessary data, and then sets parameters. On the contrary, the incremental learner adjust parametric values as each datum comes in. Even though the batch learner may seem to have some advantage over the incremental learner, since he has already observed all the relevant data, it seems more plausible to conceive language acquisition in real time as being incremental, as the author argues. Nevertheless, he introduces a further restriction on how that incremental learner proceeds: when a parameter changes its value from the default to the marked setting, all parameters that depend on it must revert to default (in case they were not already in this value). This restriction prevents the learner from making false moves when acquiring a very intricate set of interacting parameters.

The main goal of Yang (2010)'s study is to demonstrate that both syntactic parameters coded in UG and statistical generalisations must be invoked in order to account for the process of language acquisition. Yang distinguishes two types of language variation, related to the core and the periphery of the Language Faculty (Chomsky 1981). The variation in the core grammar is related to the fixation of parameters coded in the initial state of UG; and the variation in the periphery consists of language specific generalisations, such as idiosyncrasies and historical residues. Yang argues that these two types of possible linguistic variation appear to invoke also two distinct mechanisms of language acquisition: on the one hand, the core grammar is acquired by means of 'selection' among the different options that UG provides in the form

of parameters; on the other, variation in the periphery is learned by 'evaluation', a decision making process that determines the scope of inductive generalisations based on the input. Yang proves that both types of learning mechanisms display sensitivity to certain statistical properties of the linguistic input and shows how a system like that would work.

Regarding the acquisition of the core area of grammar, Yang draws on his Variational Learning Model (Yang 2002) to propose a way of capturing the dynamics of parameter setting, and introduces a probabilistic component to grasp the gradualness of acquisition:

(24) For an input sentence s, the child

(a) with probability P_i selects a grammar G_i,

(b) analyses s with G_i

- if successful, reward G_i by increasing P_i
- otherwise punish G_i by decreasing P_i

Therefore, learning the target grammar consists in selecting which parameters match the linguistic input. While alternative grammars may coexist during a period, the target grammar will eventually win. In fact, this model allows to quantify the fitness of a specific grammar from the pool of grammars given by UG as a probability of its failure in a specific linguistic environment.

According to the variational model, the time course of parameter setting need not be uniform: the more frequent the relevant data for setting a specific parameter, the faster its valuation (we will come back to this notion of "relevant data" in the next section of this chapter). Yang exemplifies this logic with some examples, such as the acquisition of the NSP, or the pro-drop parameter, across languages.

It is a well-known fact in the literature that English young children omit more subjects than English adults (Hyams 1983, Valian 1991), which can be interpreted as if they take longer than other learners (for instance, those

surrounded by typical null subject languages) in the correct fixation of the pro-drop parameter. Yang argues that this fact is due to the low presence in the input of what he considers relevant data for children to set the parameter, contrarily to the higher presence in the input to set this parameter in contexts where the target grammar is Italian or Chinese.

For the author, the relevant data is the data that can unambiguously differentiate the grammar of the three languages he considers: English, Italian and Chinese. Below there are a) the data that can unambiguously differentiate the three of them, b) the frequency percentages of these data in child directed speech, and c) the values of two parameters. Yang argues that it is not enough with the pro-drop parameter to differentiate the grammars of these three languages, given that Chinese is radical pro-drop, or topic-drop; hence he includes the value for the topic-drop as well as the value for the pro-drop parameter:

(25) Parameter settings and unambiguous data considered by Yang (2010):

(a) Chinese : Null objects (11.6), +topic-drop, -pro-drop.

(b) Italian : Null subjects in object wh-questions (10), -topic-drop, +pro-drop.

(c) English : Subject expletives (1.2), -topic-drop, -pro-drop.

Chinese is a topic-drop language, that is, it can omit constituents in the topic position. Nevertheless, there is a constraint on how this operation takes place in relation with subject omission (Yang 2002): subject drop is possible only if an adjunct is topicalised (as in (26)), thereby it is not possible if an argument is in the topic position (as in (27)). In the following examples, the *e* stands for the old discourse topic, the subject *John*:

(26) Mingtian, *e* guiji hui xiayu
Tomorrow, *e* estimate will rain
'It is tomorrow that John believes it will rain.'

(27) *Bill, *e* renwei shi jiandie
Bill, *e* believe is spy
'It is Bill that John believes is a spy.'
(Yang 2010, 1165)

Given that, we can understand why the data considered by Yang in (25) can unambiguosly point to one of the three languages. First, the relevant data for the setting of Chinese is the omission of objects, since objects can be dropped in Chinese but neither in Italian nor in English. Second, the relevant data for Italian is the drop of subjects in object wh-questions, because both Chinese and Italian can drop subjects, but only Italian can drop them when an argument, like an object, is topicalised, as in a wh-question. At this controversial point, Yang follows Chomsky (1977) in considering an object wh-question the counterpart of the topicalised sentence in (27). Sentences (28) and (29) would be the counterpart of (27), whereas the sentence in (30) would be the counterpart of (26). Again, *e* stands for omitted subjects:

(28) Chi *e* ha baciato?
Who *e* have.3sg kissed
'Who has s/he kissed?'

(29) Chi e_1 credi che e_2 ami?
Who e_1 think.2sg that e_2 love.3sg
'Whom do you think s/he loves?'

(30) Dove hai *e* visto Maria?
Where have.2sg *e* seen Maria
'Where have you seen Maria?'
(Yang 2010, 1165)

Third, the relevant data for English is the appearance of subject expletives since English is the only language of the three that has this kind of element. Yang argues in the paper, on the basis of the early acquisition of the raising of finite verbs in French, that something around the 7 percent of frequency in child directed speech can be considered a benchmark for early parameter setting. Given that result, it is predicted that both Chinese and Italian children will acquire this parameter early on, since the amount of relevant data

in child directed speech is greater than 7 percent, 11.6 and 10, respectively. However, it will take longer for English learners, since they just receive the 1.2 percent of the relevant data in child directed speech. Yang argues that the finding of these frequency effects in parameter setting strengthens the argument of a parametrised UG and the conception of language acquisition as a parameter setting process.

As for the acquisition of the periphery, the other component of grammar the author assumes, he develops a mathematical model, the productivity model (Yang 2005), that provides the conditions under which a rule becomes productive or may be considered a mere exception. Because the purpose of this sections very brief review is to outline the very basic assumptions of some learnability models in order to show how they neglect Greenberg's problem, we will not review the productivity model here. Suffice it to say that this is a model about morphological learning and processing of what can be considered the periphery of the Faculty of Language that assumes that the organization of morphology is governed by the Elsewhere Condition, and that it is a model that is able to predict how many exceptions a productive rule can tolerate.

Once this very brief review has been done and some of the most important features of the four models have been highlighted, let us now turn to the actual properties of languages of the world. In particular, I will point out three very well-known linguistic facts and I will relate them to three problems shared by some of the models reviewed in this section.

The first fact I want to focus on can be stated as follows:

(31) *Fact 1*
Natural languages exhibit 'mixed' patterns.[14]

The statement in (31) points to the fact that even accepting that it can be relevant for linguistic theory to propose wide range points of variation with associated grammatical properties, or (macro-)parameters, in order to

[14]The word 'mixed' is in quotations to indicate that is not the case that certain languages depart from a 'pure' pattern in any sense, they just look like what we may call 'mixed' merely from an observer's point of view.

differentiate big types of languages, what it is found in real linguistic samples is:

- that languages exhibit some of the properties related to that point of variation, but only in few cases all of them, and/or

- that languages exhibit the given pattern in some context or in some category, but not always in all of them.

We can illustrate both cases with the Polysynthesis Parameter we mentioned in (3). As Trask (2002) argues in his review of Baker's *The Atoms of Language*, one of the main problems of classical parametric theory is precisely the persistence of mixed patterns across languages. He takes the case of Basque to illustrate it. Whereas Basque can be considered a polysynthetic language inasmuch as it shows some of the cluster properties associated to the parameter, like subject and object agreement on the inflected verb, inflections obligatory and fixed in position, free word order, etc., it lacks noun incorporation. Therefore, Basque illustrates the case where a language exhibits some of the properties related to the parameter, but not all of them.

To exemplify the other case, where only some categories exhibit the relevant pattern, we can consider an instance of clitic doubling in Spanish. In certain dialects of Spanish, the dative clitic is mandatory, whereas the associated prepositional phrase can be omitted and appears in different positions, depending on pragmatic factors. In these varieties, the dative clitic behaves thus like an agreement affix of a polysynthetic language, as argued for instance in Kayne (2005b):

(32) Le pedí que viniera
Clitic.DAT.3sg asked that come.SUBJUNC.3sg
'(I) asked him/her to come.'

(33) Le pedí a Juan que viniera
Clitic.DAT.3sg asked to John that come.SUBJUNC.3sg
'(I) asked him/her John to come.'

(34) A Juan le pedí que viniera
To John clitic.DAT.3sg asked that come.SUBJUNC.3sg
'To John (I) asked him/her to come.'

(35) *Pedí a Juan que viniera
Asked to John that come.SUBJUNC.3sg
'(I) asked John to come.'

Although nobody would say that Spanish is a polysynthetic language, we can observe in these examples that the dative clitic behaves as an obligatory affix of a polysynthetic language in some varieties of Spanish. Nevertheless, other clitics, such as accusative clitics, do not behave like the dative marker in this Spanish variety and, accordingly, they only appear when the full phrase is omitted.

It is clear from these two illustrations, therefore, that a theory which assumes (macro-)parameters, or points of variation codified in UG, will not be able to account neither for patterns of linguistic variation among the languages of the world, nor for their acquisition. Nevertheless, all the learnability models that we have reviewed in this chapter define learning algorithms that operate with these types of parameters that differentiate big types of languages, like parameters regarding word order in Gibson and Wexler (1994) and Niyogi and Berwick (1996), or the pro-drop parameter in Yang (2010). In his study, Dresher (1999) is concerned with the valuation of phonological parameters, but he is explicit in claiming that the cue-based model can be applied to all areas of grammar, syntax included. Since he appeals to a UG with articulated parameters and cues, I understand that when he explains that his model can be applied to syntax, he intends to address this objective by means of syntactic parameters coded in UG (like *a priori* codified parameters in the other reviewed models) with associated cues. If my interpretation is correct, then the four models face the following problem regarding fact 1 in (31):

(36) *Mixed Patterns Problem*
Models assuming (macro-)parameters, with associated clustering properties, that apply to the whole language as a single unit or block

> cannot successfully be employed to deal with real natural languages, which very often show mixed patterns.

Actually, Trask (2002) points out another problem related to the conception of parameters that apply to whole languages as single units: the question of whether linguistic change is abrupt or not. Trask argues that a parametric view of grammatical change implies that changes must be abrupt, whereas historical linguistics has proved that grammatical change is almost always slow, gradual and incremental. I will come back to this point in the next section, where I will argue, following Lighfoot and Roberts, that grammatical change is abrupt inasmuch as it relies on a particular type of valuation, whereas change spread can be gradual and incremental. Nevertheless, it is possible to highlight a potential problem related to the one in (36) that has to do, at least partially, with the abruptness of grammatical change that Trask mentions. Let us assume a UG like the one assumed in Gibson and Wexler (1994) or Niyogi and Berwick (1996). In these models, the task of the learner is to choose which of the two values of parameters codified in UG must be fixed, given a concrete input. Accordingly, grammatical change will happen when the learner fixes a value for a concrete parameter that is different from the value of the parameter that has generated the input. We take as an illustration the Polysynthesis Parameter, since it has already been discussed in some detail. This model would imply that the child who acquires a different value of the Polysynthesis Parameter would attain a completely polysynthetic language, contrary to the completely non-polysynthetic language of his parents; or the other way around, the child would acquire a purely non-polysynthetic language when it is the case that his parents' language is purely polysynthetic. Leaving aside that this is not what we find in linguistic records, albeit the rapidity of some changes, this situation would even complicate the communicative continuity between the two generations of parents and children. Consider another case, like the Head Directionality Parameter (Baker 2005), that is, a parameter governing order between heads and complements very similar to the two ordering parameters used, for example, by Gibson and Wexler (1994) and Niyogi and Berwick

(1996). Given that parameter, and the dynamics of these learnability models, a grammatical change would imply that a child places all heads before their complements when it is the case that his parents place all complements after their respective heads, or viceversa. Simple logic prevents us from taking these situations as plausible. This poses serious difficulties for the learnability studies reviewed here, which can hardly be considered plausible learnability models for the acquisition of natural languages.

The second fact about linguistic variation patterns that I want to consider is the following:

(37) *Fact 2*
Natural languages exhibit morphological variation to a great extent, and in a multitude of cases this variation is associated to specific syntactic patterns.

This fact, which may seem at first sight a truism, is sometimes ignored by learnability models in so far as they only consider syntactic parameters without paying any attention to the morphological properties of target languages. Just to give a well-known example of a syntactic parameter that correlates with some morphological feature, we can consider the broadly accepted correlation between the positive fixation of the NSP and rich agreement in verbal paradigms, as argued by Rizzi (1986), among many others. Another illustration would be the correlation between radical pro-drop phenomena and the agglutinative morphology on pronouns observed by Neeleman and Szendrői (2007), to which we will come back in chapter 5.

Among the models reviewed here, at least Gibson and Wexler (1994) and Niyogi and Berwick (1996) show the following problem related to (37), whereas Dresher (1999) and Yang (2010) do consider to some extent morphological variation, although it is completely unclear how morphological properties are linked to syntactic properties in these two models:[15]

[15]In the cue-based learning model, one possibility would be to accomplish this correlation by means of codified cues.

(38) *Morphosyntactic Correlations Problem*
Models assuming syntactic parameters codified in UG cannot directly explain why some morphological properties robustly correlate with syntactic properties.

Therefore, all things being equal, learnability models that only assume traditional parameters have no way to link syntactic phenomena codified by parameters (like the possibility of subject omission in null subject languages) with cross-linguistically relevant morphological features associated to it (like rich agreement on verbal paradigms).[16] One initial possibility is to consider these correlations an epiphenomenon. Accordingly, one could argue that syntactic parameters are codified in UG and that the task of the learner is to value them according to the input (are there subject omissions in the target language? If yes, fix positively the NSP) without the necessity of paying attention to morphological phenomena (do verbs have rich agreement on verbal paradigms?), which will be no more than the result of a concrete fixation of syntactic parameters: since you can omit the subject, you have rich agreement on the verb.

Let us consider the second logical possibility (once we have excluded a third one that would consider that robust cross-linguistic correlations between morphological features and syntactic patterns are the result of mere chance): in cases where we find cross-linguistically relevant syntactic patterns that correlate with morphological phenomena, assume that the former are the result of the later, that is, that some syntactic patterns emerge as a consequence of some morphological features in the languages of the world. This will be the hypothesis entertained in this thesis, as will be developed along with the rest of this chapter and the others.

It is not difficult to evaluate the second possibility as more plausible than the first one once we consider the type of innate endowment, or First Factor in Chomsky (2005)'s terms, that the first scenario would demand. Given the first scenario, UG would have to specify all syntactic points of variation,

[16]This correlation is valid for consistent null subject languages, in Roberts and Holmberg's terminology (Roberts and Holmberg 2010).

that in turn would correlate with morphological features. Thus, a highly overspecified UG would be demanded, which following a strong tendency in the literature cannot be considered a plausible device on evolutionary terms (Chomsky 2005, Roberts and Holmberg 2010).

According to the second possibility, syntactic patterns would emerge from morphological features in the language. Given this possibility, it is not necessary to assume that all points of syntactic variation (or parameters) are codified in UG, since variation patterns would be the emergent result of independent processes associated to the morphological assembling of the language.

Both Gibson and Wexler (1994) and Niyogi and Berwick (1996) are explicit regarding this point: for their learning algorithms to work, UG must specify all assumed parameters. Dresher (1999) is also explicit in claiming that not only all parameters should be codified in UG, but also all the relevant cues associated to each parameter, which would result in an even more overspecified UG. In his study, Yang (2010) assumes as well that the core grammar is constituted by syntactic parameters coded in UG. Indeed, it seems that, although he is not explicit about this question, the "relevant" data he considers would be also part of UG given the dynamics of his system. As has been reviewed, Yang considers as relevant the data that can unambiguously differentiate the three languages in his sample regarding subject and arguments omission: null objects in Chinese, null subjects in object wh-questions in Italian, and expletive subjects in English.

Two potential problems to this proposal arise. The first one is: how does the learner know that this is the data he has to look for in the input in order to value this concrete parameter? If this information is not specified in UG along with parameters, as in the case of codified cues, it is not clear at all how the learner would know that he has to search these particular pieces of evidence. The second problem is the following: is the relevant data constrained by the sample itself? The relevant data that unambiguously differentiate the three languages seem to be adequate for that sample, but it would change if some other language was included in it. For instance, in some so-called polysynthetic languages, argument phrases can be omitted if there is

a verbal affix associated with these arguments. Therefore, in these languages the object can be omitted. This fact will make object omission a piece of ambiguous data when differentiating Chinese from, for instance, Swuahili (or Kiswuahili), a Bantu language. Accordingly, if we add Swuahili to Yang's sample, we would have to change the considered "relevant" data, perhaps indicating that the relevant data for Swuahili is the omission of the object coindexed with an affix, since in Chinese object omission is done without any overt mark. Whatever the solution to these two particular questions is, it seems clear that the UG demanded by the four learnability models reviewed in this chapter is highly overspecified.

The third fact about languages of the world that I would want to consider has already been noted in the course of this chapter:

(39) *Fact 3*
Natural languages vastly show subset/superset relations.

Berwick (1985) actually claims that it is almost always the case, in large language samples, that subset/superset relations occur:

> When data set D_i prompts a guess of language L_i, it is always possible to find another language L_j that is compatible with D_i and that can be interposed between D_i and L_i, $L_j \subset L_i$. This leads to overgeneralisation difficulties. If L_j is the correct target language, then the procedure has erroneously guessed too large a language, and, given only positive evidence, no further examples will contradict the guess L_i (Berwick 1985, 235).

We have already pointed out that null subject languages would conform a superset with respect to non null subject languages, insofar as the former allow both subject omission and subject appearance.

According to the fact in (39), the following problem can be defined:

(40) *Subset/Superset Problem*
Some learnability models do not take into consideration languages that show subset/superset relations.

Berwick (1985) notes that there are two ways of avoiding this problem. The first one is to propose a concrete order for parameter settings that prevents subset/superset problems to appear, as well as default states for parameters corresponding to their subset values. In the case of the NSP, for instance, this means that the default value would be a negative setting of the parameter, which will only change to the positive value if the learner identifies subject omissions in the input. The second one is to accept that indirect negative evidence plays a role in acquisition. As has been pointed out before in this chapter, by indirect negative evidence we mean that a learning procedure is able to assume that if it has not encountered a sentence s after some finite number of examples, then it may assume that s is not a positive example at all (Berwick 1985, 238).

Gibson and Wexler (1994) and Niyogi and Berwick (1996) are explicit in claiming that they do not consider subset languages/constructions as being part of the input for their algorithms. Nevertheless, the default initial state of parameter setting assumed in Gibson and Wexler (1994) can do away with the problem in (40). Yang (2010) does not specify anything in this respect, although we can understand his definition of "relevant data" (data that unambiguously set apart languages) as a way of trying to overcome this problem. In his study, Dresher (1999) is very concerned with the order of parameters in the learning path, that prevents the problem in (40) from arising in his model.

3.3 The causal connection between problems and conditions

The two previous sections have been devoted to show a) how parametric studies are mostly focused on Greenberg's problem and how they neglect three proposed learnability conditions; and b) how learnability approaches, which are mainly concerned with Plato's problem, cannot generally be accommodated to handle three well-known facts about natural languages.

One interesting next step would be to examine how the learnability conditions, along with their respective problems, proposed in section 3.1, and the three selected facts about the languages of the world mentioned in section 3.2, together with their associated problems, can be related. Indeed, once we inspect all of them closely it turns out that there is a causal relationship between learnability conditions and linguistic facts.

The Atomicity Condition proposed in (9) states that, given learnability considerations, parameters must be atomic insofar as acquisition proceeds in a local way, in the sense that the learner analyses the data he receives utterance by utterance and fixes local properties before fixing more general properties that depend on the fixation of minor features. This constraint on how acquisition works would cause the emergence of mixed patterns of linguistic variation, what we have called Fact 1 in (31). If the learner does not have to fix macroparameters that determine the overall nature of the target language, and instead what he is able to do is to fix the value of atomic parameters spread across the language, it is clear that the valuation of these minor points of variation will result in the appearance of 'mixed' languages that do not follow only one general pattern. In sum, neglecting the Atomicity Condition in the acquisition sphere heads towards the Locality Problem, stated in (8). Regarding patterns of linguistic variation, if the Atomicity Condition is neglected and Fact 1 is not taken into consideration, the Mixed Patterns Problem in (36) arises.

The logic of the last argument can be schematised as follows:

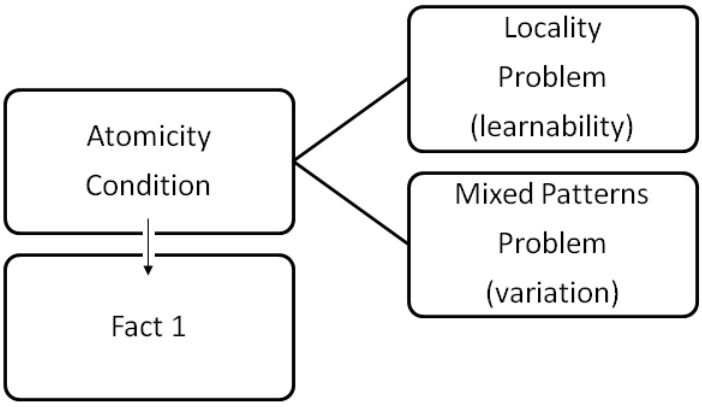


Figure 3.1: The Atomicity Condition

The Accessibility Condition in (17) dictates that unvalued points of variation must be fixed by inspecting morphophonological properties of the input that the learner is able to detect. Given this constraint on how acquisition works, it is expected that depending on which morphological properties are fixed during acquisition, the target language will show concrete syntactic patterns linked to these specific morphological features, as stated in Fact 2 in (37).[17] As I have argued, if the Accessibility Condition is neglected and parameters are defined over abstract linguistic entities, the Linking Problem, in (16), is faced. Regarding variation, if the Accessibility Condition is not considered and Fact 2 is ignored, the so-called Morphosyntactic Correlations Problem, in (38), appears.

The schema of this argument is the following:

[17]In chapter 5 it will be suggested that syntactic properties may be bootstrapped from morphological properties.

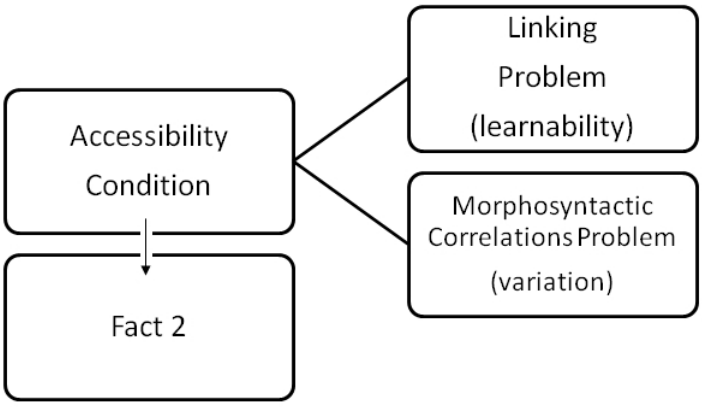


Figure 3.2: The Accessibility Condition

The Positive Evidence Condition in (19) is a learnability condition that points to the fact that the values of the parameters that must be set in order to acquire language have to be fixed by means of positive evidence present in the input of the target language. Fact 3 in (39) states that natural languages show subset and superset relationships. This third case is slightly different from the previous two cases considered in this section in the sense that the considered fact is not directly derived from the effects of the learnability condition, that is, Fact 3 is not neatly deduced from the Positive Evidence Condition. One could imagine a concrete acquisitional scenario, for instance a scenario with a sequence of ordered parameters, where subset and superset relationships do not emerge (although recall Berwick (1985)'s observation on the difficulty of finding real samples where these relations do not appear), but where the Positive Evidence Condition was relevant. For example, the values of the Polysynthesis Parameter are not in a subset relationship, but anyway the formulation of the Polysynthesis Parameter requires the use of indirect negative evidence by the learner in order to value the parameter, that is, in order to decide whether or not each argument of the verb is expressed by means of an affix on the verb. In other words, it is not the case that languages show subset/superset relations because the values of parameters must be fixed by positive evidence only; since, as we have exemplified, one can imagine a scenario where the learnability condition holds, but the possible acquistional path would be assumed not to show subset/superset relationships.

Nevertheless, I consider that it is insightful to relate both the Positive Evidence Condition and Fact 3 with the same figure as in the other cases, meaning that there is no need to be concerned about how learnability algorithms handle subset/superset relationship if a learnability condition such as the Positive Evidence Condition is active during acquisition. The Positive Evidence Condition, thus, prevents this type of phenomenon from being a problem for the learning task. As can be appreciated from figure 3.3 and as argued in this chapter, if the Positive Evidence Condition is neglected regarding acquisition, the Indetermination Problem in (18) appears; while if this learnability condition is not considered and Fact 3 is ignored, the Subset/Superset Problem in (40) arises when considering patterns of linguistic variation:

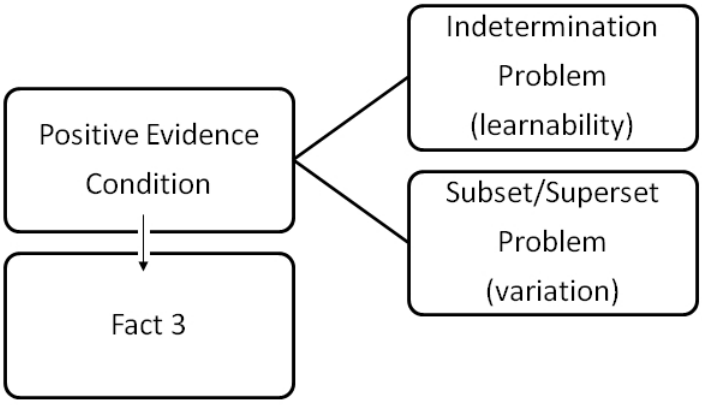


Figure 3.3: The Positive Evidence Condition

3.4 A Discovery Procedure

The goal of this section is to sketch the main features of a model for approaching the logical problem of language acquisition capable of deriving patterns of linguistic variation as emergent consequences. In this sense, this work follows the rationale behind the PP framework, whose main aim is to relate in a meaningful way acquisition and variation by means of parameters, although, as we have seen, most of the studies within the orbit of PP have been focused only on one of the two sides of the problem.

Indeed the close relation between acquisition, or Plato's Problem, and linguistic variation, or Greenberg's problem, has already been pointed out in the previous section 3.3, insofar as it has been argued that some facts about linguistic variation patterns derive from learnability conditions active during the process of language learning. Just to repeat a concrete case, it has been argued that the fact that languages show mixed patterns is a logical consequence of the fact that language acquisition is constrained by a learnability condition that states that points of variation must be atomic and have to be set locally.

Actually recall that this is the principal hypothesis of this thesis: learning mechanisms and constraints active during language acquisition shape the format of linguistic variation. In other words, some important features of the languages of the world are derived from the way in which natural languages are learned. I am arguing for this view on the bases of a) theoretical considerations and arguments developed in this chapter: learnability conditions, linguistic facts and related problems regarding both Plato's Problem and Greenberg's Problem; and b) empirical questions that will be developed in the next two chapters 4 and 5: concrete learning constraints and learning mechanisms will be independently proposed on learnability grounds, and it will be argued that they are responsible for the emergence of some patterns of variation.

Within Generative Grammar, David Lightfoot was the first to propose an explicit link between linguistic variation and language acquisition, or, more concretely, between linguistic change and acquisition. In accordance with parametric theory, Lightfoot was the first to argue that the different fixation of parametric values by children leads to linguistic change across generations of speakers. This different fixation of parametric values is triggered by the input that children receive, which may be different from the input their parents received when they previously acquired language, due to some morphological or pragmatic changes in the E-language surrounding children. Therefore, according to this view, linguistic change occurs suddenly as a result of the different fixation by children of a parameter with respect to the value of that parameter in the previous generation of speakers. According to Lightfoot

(1979), thus, languages would differ cross-linguistically and diachronically to the extent that they show different grammars, as in classical parametric theory. As Lightfoot states it:

> Each generation has to construct a grammar anew, starting from scratch. Speakers of a given grammar construct a grammar on the basis of the primary data available, i. e. the sentences in their experience. A subsequent generation constructs a grammar in the same way, but if the primary data is now slightly different the grammar hypothesised will also be different, and there is no reason why it should have to bear any close formal relation to that of the parent generation beyond the defining requirements of a theory of grammar (Lightfoot 1979, 147).

In the remainder of this thesis we will depart from this traditional parametric view in that we will follow the three proposed learnability conditions and we will take into consideration important facts about patterns of variation; however, it is important to notice the relevance of Lighfoot's work as for the link he proposed between acquisition and linguistic change.

Following some of Lightfoot's insights, it may well be that the author that has been more concerned with developing a framework to explain how acquisition is related to variation and change in a substantive way is Ian Roberts, who has emphasised the relation between them in numerous influential studies. As he argues: "while the universal structure [of language] may be either inherited or acquired, the variation must be acquired" (Roberts 2007, 7). Given this logical premise, it is almost mandatory to explore how learning mechanisms and learnability constraints influence how natural languages can possibly differ from one another.

It is important to stress that Roberts not only has related linguistic variation and language acquisition through concrete versions of parametric theory, for instance with the development of some parametric schemata, briefly introduced in 3.1.2, but he has also proposed some instances of concrete Third Factor principles that would play a role in the growth of language in the individual. I want to pay attention to two of these principles, which can

be considered principles of efficient computation: the conservatism of the learner, and the preference for simple representations, which will become especially relevant in chapter 4.

Regarding the conservatism of the learner, Roberts has extensively argued that during the process of acquisition the learner would try to acquire language in the most efficient possible way, that is, making the fewest possible different assumptions. This idea is expressed in his generalisation of the input principle, which has to do with the acquisition of marked values:

(41) *Generalisation of the input*
If acquirers assign a marked value to a head, they will assign the same value to all comparable heads (Roberts 2007, 275).

As for the simplicity of representations, Roberts has argued that learners will prefer to assign the simplest possible representation to their PLD, assuming a conception of the complexity of representations as the following:

(42) *Complexity of representations*
Given two structural representations R and R' for a substring of input text S, R is simpler than R' if R contains fewer formal features than R' ((Roberts 2007, 235), cfr. Roberts and Roussou (2003)).[18]

Actually a similar notion of simplicity, called elegance of representations, is one of the key features of the learnability model by Clark and Roberts (1993), which I will briefly summarise. It is not the case that this model is completely exempt from the problems that the other reviewed learnability models face, as for instance the impossibility of accounting for Fact 1 (31), the existence of mixed patterns. However, I consider it convenient to take into consideration some of Clark and Roberts (1993)'s insights given that the main goal of this thesis is similar to the aim of their proposal, inasmuch as their objective is to derive patterns of diachronic change from the effects of a particular algorithm of language learning.

[18]The notion of 'simplicity' is understood as a feature of linguistic representations that can be formally measured (see chapter 4), it does not imply any kind of biological simplification on evolutionary terms.

Clark and Roberts (1993) adopt the genetic algorithm approach to learnability of Clark (1992). According to this approach, the learner, who is endowed with a UG that specifies parameters that must be fixed, chooses randomly some concrete parameter settings and tries to parse the input sentences with these settings. After a round of parsing, the learner evaluates how well each setting did, and all them are ranked according to their fitness. The fittest one reproduces some candidates for the next round of parsing by combining or altering some parts of its own setting, whereas the least fit setting gets out of the ranking. Through repetitions of this procedure, parameter settings become fitter, and the fittest finally converges towards the target grammar. As the authors explain, since nothing in the approach requires the fittest grammar to be completely consistent with the one that underlies the input text, learners may arrive at final-state systems that differ from those of their parents, giving as a result diachronic change.

There are three elements that play an important role in the dynamics of this system. The first is a measure to quantify the failure of a given parameter setting to parse sentences, which the learner uses in order to rank the fittest parameter setting in a high position and to discard the least fit setting.

The second is a penalty for superset settings. The penalty for superset settings ensures that subset parameters are preferred over superset ones, since each time a superset setting is adopted, there is a penalty that makes this concrete setting less fit. Leaving aside some learnability problems posed by these dynamics, such as the fact that for this system to work the learner must be given a table where all superset and subset relations are specified in advance, the superset penalty is a way to account for the fact that subset values of parameters are actually fixed correctly by the learner during acquisition in the absence of negative evidence.

The other important element of the algorithm used by Clark and Roberts (1993) is a measure of the elegance of representations generated by a concrete parameter setting. This measure favours parameter settings that can syntactically analyse input sentences in a more elegant way, elegance being calculated in terms of simplicity. Simplicity is calculated in terms of number of nodes in syntactic trees and length of syntactic chains generated by

different parameter settings when analysing the input. Given this logic, the algorithm will rank parameter settings that allow simple syntactic representations of the input higher than parameter settings that lead to less simple syntactic representations of the input.

According to this model, the stability of parameter settings is relevant. A concrete parameter setting is stable to the extent that its expression in the input is unambiguous, and a concrete parameter setting is unstable if its expression in the input is ambiguous. For instance, parameter_a would be stable if input sentences can be successfully analysed with one value of the parameter, parameter_a^1, but not with the other value, parameter_a^2; contrarily, parameter_b would be unstable if the input can be successfully analysed both with values parameter_b^1 and parameter_b^2.

When the input data is unambiguous with respect to some parameter, it is expected that the learner will converge on the fixation of this parameter because, given the dynamics of this genetic algorithm, this parameter setting will be ranked first in the final list. However, when some parameter, for instance parameter_c, is not expressed in the input in a stable enough way, and consequently both parameter_c^1 and parameter_c^2 can successfully analyse input sentences, the learner will be forced to set parameter_c according to its internal constraints on how the acquisition algorithm proceeds: the value of parameter_c that allows the learner a simpler analysis of the input will be adopted. In other words, in cases of ambiguity in the input with respect to the expression of some parameter, the internal constraints on how language acquisition works will push the learner towards one or other of the two values. This means that internal acquisition mechanisms would drive not only how acquisition proceeds, but also how diachronic change occurs. In the authors' revealing words:

> Our argument will be that, because of various factors, the input data do not put pressure on the learner to set certain parameters to a definite value; several alternative grammars can adequately account for the input stream; the appropriate choice of grammar is underdetermined by the linguistic environment, even given the

> learner's rich internal structure. Since external pressures do not force the learner to select a particular grammar, it will turn in on itself, abandoning external pressure, and rely on its own internal structure to select from the alternatives at hand. If this is correct, then diachronic change can provide crucial information on those factors that learners rely on to select hypotheses. Since the external environment is not decisive in these cases, diachronic change reflects pure learnability considerations. Thus, diachronic change reflects what is, in a sense, "pathological" learning (Clark and Roberts 1993, 302).

Along the same lines of Clark and Roberts (1993), although without assuming the genetic algorithm, the next chapters of this thesis will be devoted to show how concrete examples of "pathological" learning occur, that is, how some proposed mechanisms and constraints active during language learning are responsible for some patterns of linguistic variation in natural languages.

It is possible to classify the learnability models reviewed in this chapter according to Chomsky (1957)'s classification on the possible relationships between a theory of linguistic structure and particular grammars. As Lightfoot explains in the introduction of the new edition of *Syntactic Structures* in 2002 (Chomsky 2002), although this was not its original purpose, Chomsky's discussion about the different ways in which linguistic theories and particular grammars can be related has also become a discussion about what the learner might be expected to do in acquisition. Following this logic, we will present the three procedures by Chomsky (1957), we will classify the reviewed learnability models as belonging to one of them, and we will argue that, in order to a) avoid difficulties faced by other models regarding Plato's problem and Greenberg's problem, and b) derive patterns of variation from learning mechanisms and constraints, a particular type of procedure should be adopted.

The three procedures are defined and represented in diagrams as in the original proposal (Chomsky 1957, 51):

(a) *Discovery Procedure.* The theory must provide a practical and mechanical method for actually constructing the grammar, given a corpus of utterances.

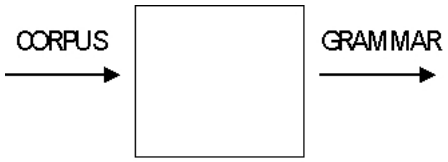


Figure 3.4: The discovery procedure

(b) *Decision Procedure.* The theory must provide a practical and mechanical method for determining whether or not a grammar proposed for a given corpus is, in fact, the best grammar of the language from which this corpus is drawn.

Figure 3.5: The decision procedure

(c) *Evaluation Procedure.* Given a corpus and given two proposed grammars G_1 and G_2, the theory must tell us which is the better grammar of the language from which the corpus is drawn.

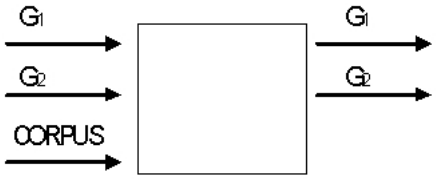


Figure 3.6: The evaluation procedure

The first two reviewed models, those proposed by Gibson and Wexler (1994) and Niyogi and Berwick (1996), can be classified as decision procedures: in both models linguistic input is analised with some concrete parameter setting; if the analysis cannot be successfully accomplished, then another parameter setting is adopted; contrarily, if the analysis can be carried out, the parameter setting is kept. In that sense the learner's task can be understood as the decision of whether or not the concrete selected grammar (or parameter setting) is suitable to analyse the input.

Lightfoot maintains in the above mentioned introduction that Dresher (1999)'s cue-based model is an instance of discovery procedure inasmuch as the task of the learner is to discover the cues in the linguistic input prior to setting parametric values. Although this is in fact the first step of the learner's task according to this model, it must be considered, at least partially, a decision procedure. This is due to the fact that all parameters and cues must be given by UG, and the learner has to decide whether or not a concrete value of a parameter can analyse the input. More concretely, given Dresher (1999)'s model the learner has to decide whether or not the given cue of the non-default value of a concrete parameter is present in the input. If the cue is present, then he turns this parameter to the non-default value. If the learner does not find the cue, he maintains the default value of this parameter. The overall procedure is then repeated with the next parameter of the ordered acquisitional path. It is in this sense that Dresher (1999)'s model must be considered a decision procedure.

Both Yang (2010)'s model and Clark and Roberts (1993)'s model are evaluation procedures, since in both cases the learning algorithms quantify the fitness of the parameter settings. This means that in both models there is a measure that quantifies the degree to which a concrete parameter setting is suitable for the analysis of the input. In Yang (2010)'s model this is achieved through the probabilistic component of the algorithm, and in the case of Clark and Roberts (1993), by means of the measurement of the failure of parameter settings in analysing the input data.

The proposal I would like to put forward in this thesis is that, in order to meet the learnability conditions proposed in 3.1 and also account for

the linguistic variation facts discussed in 3.2, a discovery procedure must be entertained. Concretely, it is proposed that a discovery procedure can better approach the answers to both Plato's problem and Greenberg's problem by considering learnability constraints derived from simplicity considerations and morphophonological mechanisms of data analysis part of the procedure:

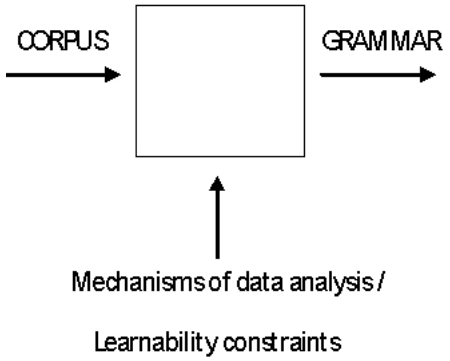


Figure 3.7: The modified discovery procedure

Learnability constraints derived from simplicity considerations and mechanisms of data analysis are assumed to be mandatory elements for the task of language acquisition; it is desirable, thus, to include them into the general picture and to derive from their effects the structure of (at least some) linguistic variation patterns observed in natural languages.

The reason behind a conception of the logical problem of language acquisition as a discovery procedure is the following: firstly, it is crucial to note that discovery procedures can avoid the learnability difficulties faced by other models, since is possible to define mechanisms of data analysis in such a way that all proposed learnability conditions are met, as we will demonstrate in chapter 5. Secondly, models that assume points of variations codified in UG, or parameters, are forced to entertain decision or evaluation procedures. Thirdly, as we have pointed out in section 3.2, the reviewed models, which propose learning algorithms that use *a priori* codified parameters, cannot be accommodated to handle some well-known facts about languages of the world.

Let us derive the claim that all models assuming concrete codified parameters in UG are necessary instances of decision or evaluation procedures. For this purpose, I follow Pinker (1979)'s observations:

(a) As Pinker (1979) demonstrates, all learning algorithms that propose a scenario where the learner considers *a priori* grammars (or parameters) in order to analyse input data are necessarily implementations of Gold's enumeration procedure (Gold 1967), which can be schematised as follows (Pinker 1979, 227):

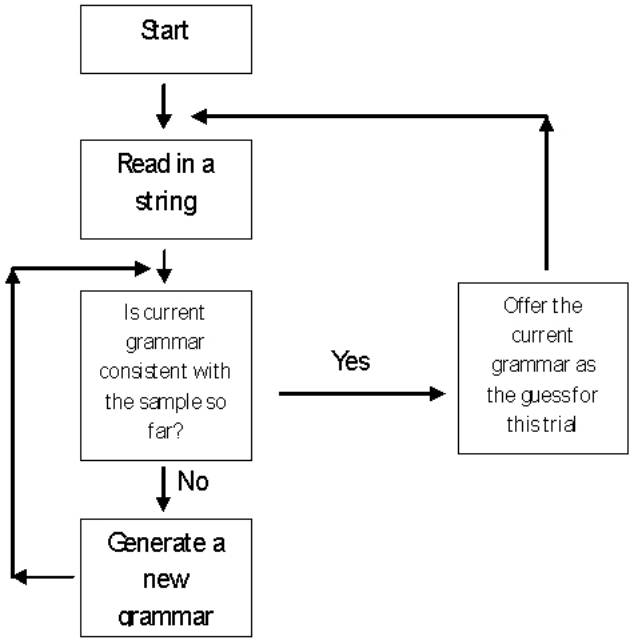


Figure 3.8: Gold's procedure

(b) Gold's theorem is unavoidably an instance of a decision procedure, or an instance of an evaluation procedure if a measure of success is included.

Therefore, it is necessary to pursue a discovery procedure in approaching language acquisition if the proposed learnability conditions and Greenberg's problem are brought into consideration. By exploring this line of research

we would dispense with *a priori* entities, that is, parameters, and we would be able to embody emergent patterns of variation (those that traditional parameters range over) in some independently motivated mechanisms.

Actually Pinker (1979) argues that Gold's procedure and all learning algorithms that can be accommodated to it, "enumerative procedures" in his terms, suffer from serious problems when facing real learning scenarios, as opposed to what he calls "heuristic procedures", which can be roughly assimilated to discovery procedures:

> They [heuristic procedures] differ from enumerative procedures in two respects. First, the grammars are not acquired and discarded as a whole, but are built up rule by rule as learning proceeds. Second, the input sentences do not just contribute to the binary decision of whether or not a grammar is consistent with the sample, but some property processed by sample sentences is used as a hint, guiding the process of rule construction. Thus heuristic language learning procedures are *prima facie* candidates for theories of human language acquisition. They acquire language piecemeal, as children do (Brown 1973), and they have the potential for doing so in a reasonable amount of time, drawing their power from the exploitation of detailed properties of the sample sentences instead of the exhaustive enumeration of a class of grammars (Pinker 1979, 234-235).

As noted by Pinker (1979), heuristic models, or discovery procedures, are much more plausible to be used by theories addressing human language acquisition than enumerative procedures from a psychological point of view.

Contrarily, on of the major advantages of enumerative models or models subsumed in Gold's general procedure is that they can be computationally modeled to guarantee the final success in identifying the target grammar, although the artificial programs designed with such an objective usually use an amount of time and memory totally implausible for a supposedly human learner.

As Pinker (1979) explains, this is why some researchers have tried to implement, with unequal results, heuristic models as computer programs to check how effective they are in concrete computational environments. In the present thesis we are not going to introduce computational machinery in order to prove how the entertained discovery procedure could work; this issue is left for future research. Suffice it to say that given the psychological implausibility of enumerative models, it is desirable to pursue another way of approaching language acquisition which can actually be accommodated to real learning scenarios, although this new approach may involve difficulties in being computationally implemented.

Apart from that, heuristic models for language learning "commit the learner to assumptions not only about the target languages, but about the sentences that find their way into the sample" (Pinker 1979, 235-236). In other words, whereas enumerative procedures endow the learner with prespecified grammars or parameters in order to analyse input sentences, heuristic models impose constraints on how the learner is going to analyse the linguistic input he will receive. As Pinker (1979) argues, this issue is completely consistent with the results obtained by Hamburger and colleagues using their mathematical model on the acquisition of transformational grammars (Hamburger and Wexler 1975, Wexler and Hamburger 1975, Culicover and Wexler 1977), since they demonstrate that learnability considerations necessarily imply strong innate constraints on the child's learning mechanisms.

In this thesis we will make explicit some of these constraints appealing to morphophonological mechanisms of data analysis from where the learner will be able to extract morphosyntactic phenomena as an heuristic strategy (see chapter 5).

Indeed Chomsky (1957) argues that for a discovery procedure to be successful, syntactic entities should be grounded on morphological units, morphological units should be grounded on phonological ones, and these on phonological and prosodic elements, meaning that for a true discovery procedure to work, different elements of grammar should be conceived as interdependent, not as completely independent elements of distinct levels of representation. What we argue for in this chapter is precisely something

very similar to Chomsky's observation: in order to answer both Plato's problem and Greenberg's problem, a) the acquisition of abstract linguistic entities should be grounded on the acquisition of more accessible properties in the linguistic input, b) *a priori* points of variation codified in UG must be dispensed with, c) learnability constraints and morphophonological mechanisms of data analysis independently motivated must be considered, and d) it should be investigated whether or not general patterns of morphosyntactic variation can be derived from the effects of learnability constraints and morphophonological mechanisms of data analysis.

Following this reasoning, in chapter 4 a learning constraint, derived from simplicity considerations, will be proposed, and I will explore how some morphological patterns in natural languages, in particular, their agglutinative or fusional nature, can be derived from this constraint. In chapter 5 I will explore how a proposed mechanism of morphophonological data analysis can be related to bootstrapping mechanisms to acquire some high-order syntactic patterns.

Before pointing out some consequences of this kind of approach, it is worth noting that this is fully compatible with UH mentioned in chapter 2, repeated here:

(43) *Uniformity Hypothesis*
In the absence of compelling evidence to the contrary, assume languages to be uniform, with variety restricted to easily detectable properties of utterances (Chomsky 2001, 2).

If we assume, according to the UH, that syntax (notably, merge) is the invariant computational device of the Faculty of Language and that linguistic variation is restricted to detectable properties of utterances, the argument that languages differ as to how they "morphologicalise" linguistic content, derived from the way in which the learner analyses his input, appears as a natural result.

Let us introduce some consequences of adopting a proposal such as the one we are developing:

(a) As has been already pointed out, we can define data analysers in a way that the three learnability conditions proposed in 3.1 are met, thus avoiding both learnability problems related to conditions and also accounting for the linguistic variation facts considered in 3.2. This point will be extensively developed in chapter 5.

(b) Considering data analysers and learnability constraints derived from Third Factors principles, it is possible to capture one of the most striking facts when dealing with linguistic variation and change: it is the case that some linguistic patterns tend to change in a predictable manner or, in other words, there is a pathway or drift in language change. Following the rationale in Clark and Roberts (1993), it is possible to relate the paths of this drift with the effects of the internal learning restrictions operating during language acquisition. In chapter 4 we will argue, following Roberts (2007), that there is a drift towards simple representations (which does not entail, of course, that linguistic change towards the opposite direction is banned, see again Roberts (2007) on this point).

(c) It is possible to maintain the three predictive areas of parametric theory, namely: acquisition, variation and change. According to parametric theory, the predictions made in these three areas are interdependent inasmuch as they all rely on the fixation of parametric values. Within the proposed model, we can also maintain the three predictive areas. In this thesis we are mostly concerned with acquisition and linguistic variation, but our main results can also be applied to the study of language change as well, as will become clear in the following chapters, where we will deal with some patterns of diachronic change.

(d) Since the linguistic entities that the mechanisms of data analysis work with are discrete, the abruptness of language change can be maintained. In this point, this proposal follows the insights of parametric theory in claiming that, despite the speed of the diffusion of concrete linguistic

changes, the actual change must be sudden in that it requires the fixation of some open value during learning (or a specific analysis of the input, as we will see in chapter 4).

There are also other considerations that favour the adoption of this proposal. First, if patterns of variation can be attributed to the effects of Third Factor mechanisms and data analysers, the problem noted by Niyogi and Berwick (1996) about the impossibility to converge with some parametric algorithms is avoided, since we just dispense with these types of algorithms that consider *a priori* codified parameters. Second, we will not face a severe problem encountered by parametric algorithms that randomly chose a parameter setting in the parameter space to begin the task of acquisition, like models by Gibson and Wexler (1994), Niyogi and Berwick (1996) or Yang (2010): it is a logical possibility that the first setting randomly chosen will correspond to the actual target grammar, and thus the learning process would be forced to end at the first step.[19] Third, if the logical problem of language acquisition is approached postulating a discovery procedure instead of enumerative procedures that require *a priori* codified parameters in order to work, the importance of the development of language in the individual is emphasised, avoiding thus the usual criticism against nativist perspectives, namely, their apparent negligence of biological development (Gopnik 1996).

Morphological analysis of the input is an unavoidable step in the process of language acquisition: the learner is supposed to go through it independently of what is assumed to come later. In the next chapters of this thesis we will try to show that it is possible to attribute to that stage more than what was traditionally assumed.

3.5 Conclusions

In this chapter we have shown how concrete proposals within parametric theory have been concerned with only one of the two problems parameters have to face with: Plato's problem, or how natural languages are learned,

[19] I thank Jordi Fortuny for this observation.

and Greenberg's problem, or what is the format of possible variation among languages of the world.

On one hand, both macroparametric and microparametric proposals have been mainly focused on deriving patterns of cross-linguistic variation and have left learnability considerations aside. In particular, we have argued that, in general, most parametric studies do not meet three proposed learnability conditions: the Atomicity Condition, the Accessibility Condition and the Positive Evidence Condition.

On the other hand, we have argued that models that apply insights from parametric theory to develop learning algorithms for language acquisition face serious difficulties when some well-known facts about patterns of linguistic variation are taken into consideration. To build this argument, we have paid attention to three linguistic facts, namely, that languages show mixed patterns, that morphological features correlate with syntactic ones, and that languages show subset/superset relations.

It has been argued that far from being independent, the proposed learnability conditions are indeed responsible for the appearance of the chosen facts about linguistic variation. Actually this is the main hypothesis entertained in this chapter and in the rest of this thesis: at least some patterns of variation among languages are derived from the way language acquisition proceeds.

In order to develop this idea, whose main roots are in Ian Roberts' studies, we have proposed a particular instance of a discovery procedure, following Chomsky (1957)'s tripartite classification, which incorporates independently motivated elements that are by assumption at work in the growth of language in the individual: Third Factor mechanisms, such as learnability restrictions derived from simplicity considerations, and morphophonological mechanisms of data analysis.

The last objective of such a project is to show that elements of grammar located in different levels of representation are interdependent, in order to propose a grounded learning procedure. This is a vast and ambitious project which will not be totally developed in this thesis; nevertheless, I hope I have shown, on the basis of theoretical arguments, why such an approach

is necessary to be entertained in order to answer both Plato's problem and Greenberg's problem. In what follows I will highlight some other advantages of adopting this research agenda by showing how concrete empirical progress can be obtained. In chapter 4, I will propose a constraint active during language learning derived from simplicity considerations which can be used to predict fusional and agglutinative patterns in languages of the world. In chapter 5, I will propose a morphophonological mechanism of data analysis capable of deriving high-order syntactic patterns by means of bootstrapping mechanisms.

4

The Minimise Paradigms Constraint

The goal of this chapter is twofold: first, a learning constraint derived from simplicity considerations is argued to be active during the process of language acquisition, specifically, during the acquisition of inflectional systems. For this purpose, I assume Pinker (1984)'s model on the acquisition of morphological paradigms, although I am going to introduce some modifications to the model in order to include in the system the learning restriction that will be proposed. Second, it is argued that some patterns of linguistic variation, particularly the fusional and agglutinative nature of affixal morphs, are derived from the effects of successive analyses carried out by learners using that constraint when acquiring morphological paradigms.

This chapter is structured in five main sections. In section 4.1 Pinker (1984)'s model on the acquisition of inflectional systems and paradigmatic relationships is presented. In section 4.2 the Minimise Paradigms Constraint (MPC), the restriction derived from the Third Factor component on how the learner builds up morphological paradigms, is proposed. In section 4.3 it is argued that the degree of fusion (Sapir 1921) in verbal morphology of natural languages is (at least partially) derived from the effects of the MPC. In section 4.4 this proposal is illustrated both with fusional and agglutinative morphological patterns: in 4.4.1 the higher fusionalisation of Romance verbal morphology with respect to Latin is explained as a result of the effects

of the MPC; in 4.4.2 the agglutinative nature of morphological verbal pieces in Chichewa and in Bantu languages in general is also argued to be a consequence of the MPC; in 4.4.3 other case studies are considered for further research. Section 4.5 is the conclusion of this chapter.

4.1 Pinker (1984)'s model on the acquisition of paradigms

In this section I will present some of the basic features of Pinker (1984)'s chapter "Inflection" in *Language Learnability and Language Development*. The goal is to briefly summarise the main characteristics of this model focusing on those aspects that will become relevant for our purposes in the rest of the present chapter regarding the acquisition of morphological paradigms. I will not assume the tenets of *Lexical Functional Grammar*, as Pinker does, although I will maintain all general principles and procedures of the original proposal.[1]

The main aim of Pinker's chapter "Inflection" is to account for the acquisition of inflectional systems, that is, for the acquisition of affixes and closed-class morphemes encoding case, number, gender, person, tense, aspect and alike categories.

After reviewing an unsuccessful previous model (Pinker 1982) that assumed that the learner had already segmented the affixes from the stems before beginning with the acquisition of the inflectional system, Pinker adduces that many of the developmental phenomena related to the acquisition

[1] As will become clear in this chapter, there are compelling arguments in favour of the use of "morphological paradigms" by the learner when acquiring his target language. It is not the goal of this chapter to review the arguments given by other theories against the use of paradigms in linguistic explanations, such as those theories that consider that paradigms are neither primitives nor elements which can be referred to by statements of grammar. Nevertheless it is relevant to note that even these kinds of theories, as for instance the *Distributed Morphology* framework (Halle and Marantz (1993, 1994) and much subsequent work), assume that "certain regularities obtaining over paradigms may result from constraints operating during language acquisition" (Rolf Noyer, Distributed Morphology web page).

of affixes, such as the speed and stability of their acquisition, the discovery of zero morphemes, involve variables sensitive to their segmentation.[2]

In order to develop another model for the acquisition of affixes that includes the segmentation of affixes and stems as a fundamental step, Pinker incorporates the traditional notion of *paradigm*. Indeed, as Pinker notes, the segmentation of affixes "is a nontrivial step that cannot be accomplished by an examination of individual inflected words" (Pinker 1984, 172), and the inclusion of morphological paradigms in the theory can be used as an index of comparison of related forms, as will become clear later.

Before summarising the basic features of the model, let us provide a formal definition of *paradigm* and make clear some other related concepts.

In the literature, the term *paradigm* has different, though related, meanings. We adhere to the following definition in the present chapter:

(1) *Paradigm*
Matrix of inflectional realizations appropriate to a given inflectional class (Carstairs-McCarthy 2001, 323).

As in Carstairs-McCarthy (2001), we take an *inflectional class* to be a set of lexemes which share a paradigm and whose word forms are alike with respect to the realisation of the morphosyntactic properties in every cell. In general, the term inflectional class is replaced by *declensional class* for nouns and adjectives, and by *conjugational class* for verbs.

A consequence of using paradigms to represent inflectional systems is that, instead of classifying declensional information solely by appending grammatical features to the lexical entries for each affix, the grammatical information itself can also serve as an indexing system, under which particular affixes are listed (Pinker 1984, 174). In the paradigm below, which corresponds to Latin first declension, affixes are ordered with respect to the grammatical information of number (columns) and case (rows):

[2]The *stem* can either be the contiguous strings of segments in languages with prefixing or suffixing, or the triconsonantal root in Semitic languages (Pinker 1984, 188).

	Singular	Plural
Nominative	a	ae
Accusative	am	as
Genitive	ae	arum
Dative	ae	is
Ablative	a	is

Table 4.1: Example of paradigm: Latin first declension

Number and case are considered the two *dimensions* of the paradigm in 4.1, each of its possible values like nominative or singular are *levels* of the respective dimensions, and the affixes are listed within *cells* representing a conjunction of levels of different dimensions (Pinker 1984, 175).

An important constraint on affixation that can be stated easily in terms of the paradigm representation is that no complete set of grammatical feature values may be encoded by two or more distinct morphs. This can be translated into the constraint that no cell in a paradigm may be filled with more that one affix, or the *Unique Entry Principle* (UEP), in Pinker's terminology (Pinker 1984, 177). This constraint will be crucial for the dynamics of the model and will impel the learner to acquire certain morphological representations. Despite that, there are counterexamples to this constraint in a great number of languages, like, for example, the alternation between the two forms of the imperfect and the pluperfect tenses in subjunctive mood in Spanish: "comi*era*"-"comi*ese*" (*I would eat*), and "hubi*era* comido"-"hubi*ese* comido" (*I would have eaten*).

Thus the notion of paradigm or matrix representation can be used to represent adult knowledge about sets of related affixes. As Pinker notes, there are theoretical and empirical reasons to believe that both general and word-specific paradigms are necessary in adult grammar (Pinker 1984, 175-176). Therefore, under this view, the task of the learner is to build up the specific and general paradigms of his target language in order to acquire the inflectional system.[3]

[3]Pinker refers to "words" in some of his explanations. However, we do not commit

In what follows, Pinker (1984)'s model is summarised in three blocks: the formation of word-specific paradigms, the formation of general paradigms and the discovery of arbitrary classes. Even though some of the procedures are simplified, the overall mechanics of the model is maintained.

(a) *The formation of word-specific paradigms.*

The goal of the learner in this first stage of the model, where segmentation between stems and affixes is not done yet, is to build word-specific paradigms of his target language. Given that the developmental evidence supports the progression from specific to general paradigms (Pinker 1984, 180), it is natural to start studying the acquisition of inflectional systems at this step.

The proposed procedure for the formation of word-specific paradigms is the following: the learner chooses a linguistically relevant feature, or morpheme, in the inferred sentence meaning and creates a one-dimension paradigm. For the paradigm below, the learner has chosen the features of 'nominative', 'accusative' and 'dative' from the abstract dimension of case, and has created the paradigm for the concrete word 'ti' in a hypothetical language:

Nom	Acc	Dat
tia	tib	tic

Table 4.2: Word-specific paradigm with one dimension (ti)

Following the formation of the word-specific paradigm for 'ti', the learner notes that in the input there is more than one form to encode the features 'nominative', 'accusative' and 'dative' and includes all these possibilities in the respective cells in the paradigm:

ourselves to the assumption that the notion of "lexical word" has a primitive status neither in linguistic theory nor in language acquisition.

Nom	Acc	Dat
tia/tid	tib/tie	tic/tif

Table 4.3: Word-specific paradigm in construction (ti)

This constitutes a systematic violation of the UEP, which guides the learner towards the postulation of a new dimension in the paradigm. Then, when the lexical entry for a word that is already in the paradigm is given new hypothesised information involving different features, say 'singular' and 'plural', this information is encoded as a new dimension in the paradigm. At this point, our 'ti' word would have a specific paradigm with five different levels:

	Nom	Acc	Dat
Sg	tia	tib	tic
Pl	tid	tie	tif

Table 4.4: Word-specific paradigm with two dimensions (ti)

It is important to note that the paradigm resulting from this procedure is not set in stone: a hypothesised affix will be retained in a word-specific paradigm or will be replaced by another one depending on the input evidence that the learner encounters. This means that the forms that enter into the paradigm are in competition with new possible forms hypothesised by the learner for the same level.

Pinker notes in a footnote (Pinker 1984, footnote 12, 386) that the proposed procedure to create word-specific paradigms could erroneously treat free closed-class morphs as if they were bound morphs. For instance, the learner could analyse the string "thedog" as if the determiner "the" was a definiteness-marking prefix and could mistakenly build the following paradigm:

Definiteness
thedog

Table 4.5: Word-specific paradigm for a definiteness-marker

Pinker proposes, then, that the procedure could be modified so as to categorise an affix as a free morph if another free morph (e. g., an open-class word as in *the big dog*) was heard to intervene between the supposed "affix" and the stem. The fact that the procedure can be modified in order to keep track of syntagmatic relationships will be crucial for us, as will become clear in the next section of this chapter.

Once a word-specific paradigm is built, the same procedure is applied to different words in order to create word-specific paradigms for them. The learner would apply, for instance, the same procedure to the hypothetical words 'to' and 'pu'. First, he would create a one-dimension paradigm after choosing some level distinctions, like 'nominative', 'accusative' and 'dative'. In our invented language, 'to' would belong to the same inflectional class as 'ti', but 'pu' would belong to a different one:

Nom	Acc	Dat
toa	tob	toc

Table 4.6: Word-specific paradigm with one dimension (to)

Nom	Acc	Dat
puz	puy	pux

Table 4.7: Word-specific paradigm with one dimension (pu)

After that, the learner would continue with the same procedure and would enlarge the word-specific paradigms with more dimensions:

	Nom	Acc	Dat
Sg	toa	tob	toc
Pl	tod	toe	tof

Table 4.8: Word-specific paradigm with two dimensions (to)

	Nom	Acc	Dat
Sg	puz	puy	pux
Pl	puw	puv	put

Table 4.9: Word-specific paradigm with two dimensions (pu)

The question that arises is how the learner hypothesises abstract features, as 'nominative' or 'plural', in order to create word-specific paradigms. Pinker suggests, following developmental studies on language acquisition (Pinker 1984, 182-183), that some features are hypothesised early due to their semantic and contextual salience in the speech act where the word is uttered. For instance, the category of number applied to nouns usually has real correlates in the communicative situation that the learner can appreciate. Case morphemes, as 'nominative' or 'accusative', can also be part of these types of semantically salient features that are hypothesised early: the learner would infer that certain affixes appear related to agent actors and others, to patient referents. Then, he would create a dimension in the paradigm with different levels encoding that information. Given this logic, Pinker suggests that less semantically accessible and arbitrary features would be later hypothesised by the learner. This idea of semantic bootstrapping would explain why in the beginning different learners consistently fail to mark some distinctions but not others (Pinker 1984, 183). We will come back to the acquisition of less salient features in the last part "The discovery of arbitrary classes".

As Pinker notes, another trait that has to be taken into consideration regarding how the learner hypothesises new grammatical features is whether or not they show some phonological correlation, since it has been proved that features with phonological correlations are acquired earlier and more robustly (Slobin 1984). Pinker introduces the variable of the correlation of phonological features with morphological categories in the general procedure to acquire inflectional systems, although I will not review this last point since it does not affect the general argumentation.

In brief, the result of this first stage in the acquisition of the inflectional system is the construction of several word-specific paradigms, one for each specific chosen word.[4]

(b) *From word-specific to general paradigms.*

Once the learner has created word-specific paradigms, he has to discover what each affix encodes when it inflects any stem and, consequently, segmentation between stems and affixes must be accomplished.

For segmentation to take place, the learner has to find the phonetic material in common among all the cells in a word-specific paradigm and classify that material as "stem".[5] In our examples above, the learner would be able to discover that 'ti', 'to' and 'pu' are stems in his input language because this information is present in all the cells of their respective word-specific paradigms.

Subsequently the learner can create general paradigms with the material that remains in the cells and, thus, he can discover the inflectional system of his target language. The general paradigm for stems such as

[4]Another interesting consequence of the dynamics of this model is that, by first hypothesising semantic equations for entire verbs and only later deciding whether this information belongs to the stem or to the affixes, the learner can learn if the language encodes semantic distinctions in affixes or using verbal stems alternations (Pinker 1984, 180).

[5]As Pinker notes "I use the phrase 'phonetic material in common' as a placeholder for more precise notions to be taken from a theory of phonology" (Pinker 1984, 188).

'ti' and 'to' would be the the one in 4.10, and the paradigm for stems as 'pu' would be the one in 4.11:

	Nom	Acc	Dat
Sg	a	b	c
Pl	d	e	f

Table 4.10: General Paradigm with two dimensions (ti, to)

	Nom	Acc	Dat
Sg	z	y	x
Pl	w	v	t

Table 4.11: General Paradigm with two dimensions (pu)

This methodology is useful not only to discover overt affixes, but also to discover "zero morphemes", that is, to discover covert morphemes. Let us consider a word-specific paradigm like the one in 4.12:

	Nom	Acc	Dat
Sg	taka	takate	takati
Pl	takama	takami	takamo

Table 4.12: Word-specific paradigm with covert morphemes

Once the learner has created the general paradigm and has extracted the phonetic material in common that corresponds to the stem, the general paradigm generated is the following:

	Nom	Acc	Dat
Sg	⊘	te	ti
Pl	ma	mi	mo

Table 4.13: Word-specific paradigm with covert morphemes

Given that a zero entry is left in the cell for 'nominative, singular', the learner can discover the existence of a covert affix.

As can be observed, the procedure for creating general paradigms allows us to state straightforward mechanisms for the segmentation of affixes and stems, which constitutes an important benefit of the model.

Once the general paradigm has been formed, the learner can apply its information to incomplete word-specific paradigms to fill in empty cells, which may result in overregularization cases (Pinker 1984, 192). Actually, Pinker proposes that the coordination between word-specific paradigms and general paradigms is necessary not only to fill in word-specific paradigms with empty cells, but also to check the consistency of the language's full affixing regularities once paradigms have already been filled in. This process reinforces the content of general paradigms during acquisition. As Pinker explains "this process of retroactive constraint reinforcement is a natural extension of the use of general paradigms to alter word-specific ones requiring only the additional assumption that general paradigms are always invoked when an inflected form is accessed, not just when a desired form is absent" (Pinker 1984, 196-197).

The procedure to create general paradigms as stated so far treats fusional and agglutinative morphology identically: regardless of the morphological nature of the affixes in the paradigms, an unanalysed affix is listed in a cell as encoding the combination of feature values specified by the position of the cell in the matrix. As Pinker remarks, a procedure that treats the acquisition of agglutinative and fusional patterns alike cannot explain why there are so many consistent differences

between the acquisition of both patterns. For instance, it has been demonstrated that agglutinative patterns are much easier and quicker to be acquired by the learner than fusional morphology (Slobin 1984).

It is worth noting that Pinker does not provide any definition for the notions of "agglutinative" and "fusional" paradigms in his proposal. Indeed one can find several definitions of these concepts in the literature, some of them quite different from one another. Moreover none of these definitions, as far as I am aware, are precise and formalised, since they usually appeal to vague and undefined notions such as "word".

Since these two concepts will be very relevant in this chapter, for our purposes we adopt Bauer (2004)'s definitions, which are reproduced below:

(2) *Agglutinative pattern*
An agglutinative language is a synthetic language where the normal pattern is for each morph to realise a single semantic unit or morpheme, and each morpheme to be realised by a single morph (Bauer 2004, 15).

(3) *Fusional pattern*
A fusional language is a synthetic language in which, for some significant part of the inflectional morphology, it is not possible to isolate a morph to correspond to every semantic unit which can be distinguished in the word form (Bauer 2004, 46-47).

As Sapir already recognises, the terms "agglutinative", "fusional" or "synthetic" are relative, and a language can be "agglutinative" from one standpoint, but "fusional" from another (Sapir 1921, 128). It is because of that reason that throughout this thesis we will apply these labels to specific patterns and not to languages. Actually, Bauer illustrates both of his definitions with sentences belonging to the same language, English:

> In an English sentence like *Cats mistrusted dogs in earlier times* all the words except "in" show a basic pattern of agglutinating morphology, though the "s" on "cats" and "dogs" is pronounced differently (...) In an English sentence like *Those were stolen*, each word can be argued to show fusional morphology: "those" because it contains the ideas of 'that' and 'plural', and we cannot analyse it neatly into two parts, "were" because it contains the notions of 'be' and 'past tense' and 'plural subject', and "stolen" because it contains the notions of 'steal' and 'past participle' (Bauer 2004, 15 and 47).

Given that Pinker's procedure for generating general paradigms from word-specific paradigms treats agglutinative and fusional morphology alike, the author proposes that the procedure must be refined in order to be able to grasp the differences between the acquisition of both patterns. Accordingly Pinker proposes to split the procedure that has just been sketched, (b) *From word-specific to general paradigms*, into two procedures: once the stem is already segmented, the learner applies a procedure suitable to acquire agglutinative paradigms in a first place (I), and just if that mechanism does not work, the learner applies an alternative procedure (II). These two procedures are the following:

- I. *Creating general paradigms once the stem is segmented.* The learner chooses a dimension from a multidimensional word-specific paradigm and a level of that dimension, examines all the cells of the paradigm specified for that level, and extracts the common phonetic material. Then, he enters that material in a cell of a unidimensional general paradigm corresponding to that level. This mechanism is repeated for the other levels of that dimension to complete the paradigm. Once this is done, the learner repeats the entire process for other dimensions in order to build unidimensional paradigms for them (Pinker 1984, 189-190).

  Let us illustrate this mechanism with the agglutinative paradigm in 4.14 (Pinker 1984, 188):

	Nom	Acc
Sg	ip	iq
Pl	jp	jq

Table 4.14: Idealised agglutinative paradigm

Imagine that the learner chooses the dimension of case and the level 'nominative'. Once he has examined all the cells specified for that level, he can extract the phonetic material in common 'p'. Then he continues with the other level of the same dimension, 'accusative', and applies the same procedure to extract the common material, in this case 'q'. At this point, the learner is able to construct a unidimensional general paradigm like the one in 4.15

Nom	Acc
p	q

Table 4.15: Agglutinative suffixes for case

At this point, the learner repeats the whole process to build the other unidemensional paradigm in 4.16, which corresponds to the number dimension:

Sg	Pl
i	j

Table 4.16: Agglutinative suffixes for number

- II. *Creating general paradigms once the stem is segmented.* When the learner of a fusional pattern, like the one in table 4.17 (Pinker 1984, 188), applies the former procedure (I) and chooses a dimension and a level, he finds no phonetic material in common and, thus, he cannot create a general unidemenisonal paradigm.

	Nom	Acc
Sg	a	b
Pl	c	d

Table 4.17: Idealised fusional paradigm

In that case, the learner has to choose two dimensions simultaneously, considering in turn all four combinations of the levels of the two dimensions and searching for common material. If it does not exist, as happens in the paradigm 4.17 that we are considering, material is just extracted and placed in an n-dimensional paradigm, resembling the previous one:

	Nom	Acc
Sg	a	b
Pl	c	d

Table 4.18: Fusional suffixes for case and number

Pinker argues that the developmental data that show that agglutinative patterns are easier to acquire are explained by his model insofar as the learner must perform a more complex set of calculations and needs a major processing capacity for building fusional paradigms than for building agglutinative paradigms (Pinker 1984, 191). The reason is that the learner of an agglutinative pattern does not have to go through the second part of this process (II) where the formation of multi-dimensional paradigms is necessary.

In summary, at this stage of the procedure, once (I) and/or (II) have been carried out by the learner, segmentation between stems and affixes is accomplished, and the learner has acquired the general paradigms of his target language. On the one hand, the learner has created unidimensional paradigms for agglutinative affixes; on the other hand, the learner has created multidimensional paradigms for fusional affixes.

(c) *The discovery of arbritrary classes.*

The goal of the learner in this third block is basically the same one as in the previous two stages: he has to build word-specific and then general paradigms of the inflectional system of his target language. However, when the learner has to acquire arbitrary declensions, conjugational classes, affixes sensitive to less accessible semantic and phonological features or exceptional words, the proposed semantic bootstrapping does not help the learner to postulate the necessary morphosyntactic features to begin with the procedures. In other words, sometimes there are not (enough) semantic or phonological salient features to be used by the learner in order to begin the construction of paradigms; as for example when the learner has to acquire affixes that express arbitrary declensions or conjugations.

In these cases, the role of the UEP is essential: the learner can postulate and acquire these types of morphosyntactic features as a response to systematic violations of the UEP in his paradigmatic representations.

For instance, the learner cannot infer from the context the existence of two conjugational classes for two types of verbs if they do not correlate with phonological or semantic features. What the learner encounters is simply that the general paradigms he can build systematically contain two elements in each cell:

	Sg	Pl
1st	a/e	b/c
2nd	i/o	d/f
3rd	u/y	g/j

Table 4.19: Systematic violations of the UEP

At this point, the learner is impelled to notice that the language was making a previously unnoticed distinction, and he splits the paradigm:[6]

[6]Pinker argues that the difference in acquisition in how inflectional affixes are pos-

	Sg	Pl
1st	a	b
2nd	i	d
3rd	u	g

Table 4.20: Acquisition as a result of systematic violations of the UEP (I)

	Sg	Pl
1st	e	c
2nd	o	f
3rd	y	j

Table 4.21: Acquisition as a result of systematic violations of the UEP (II)

Indeed the role of the UEP goes beyond that, a uniqueness violation can also inform the learner about the type of missing morphosyntactic feature in his paradigm: whether a uniqueness violation occurs in a word-specific or a general paradigm is a diagnostic of whether a missing dimension is context-specific or stem-specific, respectively. Let us make this point clear. First, imagine a scenario where a systematic uniqueness violation occurs in a word-specific paradigm. For instance, imagine a learner who hears the stem X inflected with the suffix 'ma', and also with the suffix 'na':

tulated, that is, or by semantic/phonological bootstrapping or by systematic uniqueness violations, can be responsible for why some morphosyntactic features are more common cross-linguistically than others: "the difference between common inflectional features such as number, gender, and case, and rare ones such as shape and relative position in a supernatural hierarchy, would correspond to the distinction between features that are spontaneously hypothesised and features that are noticed only as a result of uniqueness violations" (Pinker 1984, 198). Although we are not going to explore this correlation here, it is interesting to note that Pinker himself proposes a relationship between mechanisms of acquisition and patterns of variation very consistent with the perspective and the assumptions of the present thesis.

Xma/Xna

Table 4.22: UEP violation in a word-specific paradigm

The reason why the learner builds a word-specific paradigm with more than one form per cell is that he hears in his input the same word, X, with different affixes depending on the communicative context where the word is uttered. In our example, suffixes 'ma' and 'na' can be the exponents of the nominative and the accusative, for instance, or exponents of the definite and the indefinite morphemes.

Second, imagine a scenario where a systematic uniqueness violation occurs in a general paradigm. The learner has built the general paradigm in 4.25 for the nominative marker after extracting stems, Y and Z, from word-specific paradigms in 4.23 and 4.24:

Nom
Yte

Table 4.23: Word-specific paradigm to illustrate a UEP violation in a general paradigm (I)

Nom
Zse

Table 4.24: Word-specific paradigm to illustrate a UEP violation in a general paradigm (II)

Nom
te/se

Table 4.25: General paradigm to illustrate a UEP violation in a general paradigm

In this case the reason why there is more than one form per cell in the general paradigm is because the stem Y has a maker for the nominative, 'te', different from the marker for the nominative that Z requires, 'se'. Thus although stems Y and Z are inflected for the same feature, they have different affixal forms. In this case, the learner can discover that the UEP violation in the general paradigm is due to factors specific to the nature of the stem. In our example, Y would belong to a different declension than Z.

Therefore, a systematic uniqueness violation in a word-specific paradigm impels the learner to find a context-specific missing feature. Nevertheless, a uniqueness violation in a general paradigm informs the learner of the necessity for looking for a stem-specific missing distinction (Pinker 1984, 199).

Moreover, the use of the UEP to hypothesise new paradigm dimensions assures the learnability of inflectional systems in the absence of negative evidence. As Pinker argues, if there was no constraint limiting a cell in the paradigm to containing one affix, the German learner would possibly build a paradigm like the one in 4.26, instead of the more restrictive in 4.27 (Pinker 1984, 201-202):

Nom	Acc	Dat
der/die/das	das/die/den	dem/der

Table 4.26: Underspecified paradigm of determiners in German

	Nom	Acc	Dat
M	der	den	dem
F	die	die	der
N	das	das	dem

Table 4.27: Fully specified paradigm of determiners in German

> As long as the learner does not use negative evidence to correct the representation in 4.26, there will be nothing to impel him to change from the underspecified paradigm in 4.26 to the more inclusive paradigm represented in 4.27, in which the choice of the article is contingent on the gender of the noun.

Before concluding this section, let us summarise the basic features of Pinker (1984)'s model presented so far and its main advantages.

The model for the acquisition of the inflectional systems of natural languages by means of morphological paradigms consists in a) a procedure to hypothesise morphosyntactic features by means of semantic bootstrapping and the correlation of phonological features, b) the formation of one-dimension and then n-dimension paradigms of related sets of affixes and words, c) a progression from word-specific to general paradigms, and d) the extensive use of the UEP to inspire changes in the paradigms already built.

One of the major benefits of the present model is that the segmentation problem during language acquisition is simplified, given that paradigmatic representations allow the learner an easier approximation to the problem of segmenting affixes and stems and to the question of discovering zero morphemes (Pinker 1984, 207).

Another advantage is related to the "formidable research problem", in Pinker's words (Pinker 1984, 168), that the learner is faced with when trying to determine which notions are encoded in his language's morphology. Actually the question of how the learner assigns semantic content to the discrete units present in his input language is one of the main concerns of the acquisition literature. Assuming morphological representations in paradigms, this question is simplified: "by entering different forms in cells constituting a row of a paradigm, the learner is implicitly 'expecting' the language to have forms for feature values other than the first one hypothesised"(Pinker 1984, 186).

This advantage is closely related to the Generalisation of the input introduced in chapter 3, repeated here:

(4) *Generalisation of the input*
If acquirers assign a marked value to a head, they will assign the same value to all comparable heads (Roberts 2007, 275).

This Third Factor principle can be accommodated to the proposal made here in the sense that, once the learner has assigned a semantic value to a form in a given cell, he will extrapolate this information to the other morphs that will enter into the same column, optimising the general process. The very formalism of paradigms, therefore, appears to be at least motivated by the dynamics of operating Third Factor principles, as the one in (4).

In the present model the important question of how arbitrary classes are acquired is addressed by appealing to the UEP, which states that there can only be one form per cell within paradigm representations. By the effect of this principle, the learner is impelled to discover some previously unnoticed morphosyntactic features.

We have not reviewed the question of syncretism in this section, but Pinker indicates that an improvement of the model with respect to previous ones is that using paradigmatic representations allows us a more natural explanation of well-known patterns of syncretism, which would be a phenomenon typically found in contiguous cells within the paradigm if one nests some dimensions within others (Pinker 1984, 178). Nevertheless, a similar sort of explanation is also found in other theories that do not make use of paradigms, which assume that syncretisms are found in contiguous nodes given a cartographical representation, as assumed for instance in Caha (2009).

4.2 Defining a learning restriction: the Minimise Paradigms Constraint

In this section we propose a learning restriction derived from the Third Factor component on how the learner builds up morphological paradigms of his target language and, consequently, we introduce a change in Pinker (1984)'s model in order to capture the effects of this constraint.

The following restriction is proposed:

(5) *Minimise Paradigms Constraint* (MPC)
The learner hypothesises just one general paradigm when affixes filled in different cells show a systematic syntagmatic relationship.

A more informal way of reformulating the MPC is stating that the learner prefers to hypothesise the fewest possible paradigms when affixes show a systematic syntagmatic distribution, a concrete pattern detected by the LAD.

Let us introduce some modifications in Pinker (1984)'s model in order to illustrate how the MPC works within the whole system. We propose that, once the stem is separated from the affixes in the first stage of the model, the procedure can keep track of syntagmatic relationships between affixes; particularly, we propose that the procedure can determine whether or not affixes left in cells once the stem is segmented show a systematic syntagmatic relationship.

Let us illustrate these dynamics with the Latin inflectional system, to which we will come back extensively in the next sections of this chapter. In Latin, the verb is formed by the stem, the temporal/aspectual/modal (TAM) marker and the personal desinence. In infectum tenses (present indicative and subjunctive, imperfect indicative and subjunctive, and imperfect future) there are two sets of personal desinences, one for the active voice and another for the passive voice:

	Singular	Plural
1	-o, -m	-mus
2	-s	-tis
3	-t	-nt

Table 4.28: Personal desinences for infectum forms in the active voice

	Singular	Plural
1	-r	-mur
2	-ris, -re	-mini
3	-tur	-ntur

Table 4.29: Personal desinences for infectum forms in the passive voice

Accordingly, infectum verbal forms in active and passive voice have the same stem and the same TAM markers but vary in the personal desinence. In the examples below, the two verbs share the stem of the verb *am-*, the TAM marker of the imperfect 'ba', and the number and person of the desinence, but differ in voice:

(6) amabat
love.TAMimp.active3sg
'He/She loved' (imp.).

(7) amabatur
love.TAMimp.passive3sg
'He/She was loved' (imp.).

Perfectum forms (pluperfect indicative and subjunctive, perfect future, and perfect indicative), however, do not form the passive by means of dedicated personal desinences, but through an analytical phrase composed of by a past participle and the verb *sum* ('to be') conjugated:

(8) Equus vulneratus est hasta
horse.NOMsg injure.pp to-be.3sg spear.ABLsg
'The horse was injured by a spear.'
(Valentí i Fiol 1995, 106)

The active voice of perfectum tenses just consists of the stem, the TAM marker and the personal desinences, which for perfectum tenses are the same as in 4.28 (except for the perfect tense, which has a dedicated paradigm).

Let us imagine a learner who is trying to acquire a system like this and forms word-specific paradigms, using Pinker (1984)'s model. Some of the

word-specific paradigms for the verb 'amo' (to love) built by the learner would be the following (infectum and perfectum forms are grouped just for the sake of clarity; remember that these are word-specific paradigms):

	Imp Ind Act	Imp Ind Pas	Imp Fut Act	Imp Fut Pas
1 sg	Amabam	Amabar	Amabo	Amabor
2 sg	Amabas	Amabaris, Amabare	Amabis	Amaberis, Amabere
3 sg	Amabat	Amabatur	Amabit	Amabitur
1 pl	Amabamus	Amabamur	Amabimus	Amabimur
2 pl	Amabatis	Amabamini	Amabitis	Amabimini
3 pl	Amabant	Amabantur	Amabunt	Amabuntur

Table 4.30: Some infectum forms of the verb 'amo' (to love)

	Pluperf Ind	Pluperf Sub	Perf Sub
1 sg	Amaveram	Amavissem	Amaverim
2 sg	Amaveras	Amavisses	Amaveris
3 sg	Amaverat	Amavisset	Amaverit
1 pl	Amaveramus	Amavissemus	Amaverimus
2 pl	Amaveratis	Amavissetis	Amaveritis
3 pl	Amaverant	Amavissent	Amaverint

Table 4.31: Some perfectum forms of the verb 'amo' (to love)

When stems are extracted, the general-paradigms under construction would be the following:

	Imp Ind Act	Imp Ind Pas	Imp Fut Act	Imp Fut Pas
1 sg	bam	bar	bo	bor
2 sg	bas	baris, bare	bis	beris, bere
3 sg	bat	batur	bit	bitur
1 pl	bamus	bamur	bimus	bimur
2 pl	batis	bamini	bitis	bimini
3 pl	bant	bantur	bunt	buntur

Table 4.32: Some infectum desinences of the verb 'amo' (to love)

	Pluperf Ind	Pluperf Sub	Perf Sub
1 sg	eram	issem	erim
2 sg	eras	isses	eris
3 sg	erat	isset	erit
1 pl	eramus	issemus	erimus
2 pl	eratis	issetis	eritis
3 pl	erant	issent	erint

Table 4.33: Some perfectum desinences of the verb 'amo' (to love)

According to Pinker (1984)'s system, having reached this stage, the learner would have to create general paradigms applying, first, the procedure to isolate "agglutinative morphemes", that is, those that are present in every cell of the same level and dimension and, then, if this mechanism does not work due to the fact that there is no phonetic material in common in all the cells of a level, the learner would apply the second procedure and would try to build multi-dimensional paradigms (we have exemplified these procedures in tables 4.14 and 4.17 in the previous section).

It is at this point where we have to modify Pinker (1984)'s model in order to introduce the effects of the MPC. We will define the concept of being in a "systematic syntagmatic relationship with a morph" and then we will develop our proposal.

We say that morphs in a paradigm show a systematic syntagmatic relationship with morphs after them when they always appear adjacently, the first is always followed by the other, and no other material can appear in between. We say that morphs do not show such a systematic syntagmatic relationship when they can appear not adjacently or one can appear without the other.[7]

In the illustration below, Morph_1 does not show a systematic syntagmatic relationship with morph_2 because morph_3 can appear between them or only with morph_1, and morph_2 can also be absent:

- Morph_1 + (Morph_3) + Morph_2
- Morph_1 + Morph_3
- Morph_1

However, in the configuration below, it is always the case that Morph_1 show a systematic syntagmatic relationship with Morph_2, since they always appear adjacently, no other morph can appear in between and Morph_1 cannot appear without Morph_2:

- Morph_1 + Morph_2
- Morph_1 + Morph_2 + Morph_3
- Morph_3 + Morph_1 + Morph_2

Just to give another example, let us focus on the configuration displayed by Spanish nouns. In most morphological theories and frameworks, the final vowel of Spanish nouns is analysed as the gender marker, as for instance in *cas-a* ('house'), *maestr-a* ('teacher') or *cuadr-o* ('picture'). In Spanish, the plural on nouns is almost always marked by means of the addition of a final *-s*: *cas-a-s*, *maestr-a-s*, *cuadr-o-s*. One could wonder, then, whether

[7]Although it is clear from the definition of what it means to be in a systematic syntagmatic relationship, it is worth emphasising that this configuration refers to overt morphology.

the gender marker and the plural marker are in a systematic syntagmatic relationship in Spanish nouns. The answer is negative: in singular contexts gender markers appear without plural markers, thus, it is not the case that the two morphs in question always appear together and one cannot appear without the other.

Therefore, according to the definition, the TAM affixes 'ba' and 'b'/'bis' in table 4.32, repeated in table 4.34, do not show a systematic syntagmatic relationship with the affixes after them because it is not the case that these TAM affixes are always followed by the same set of affixes: in one paradigm TAM markers are followed by some affixes, the active suffixes, and in other word-specific paradigm TAM markers are followed by some others, the passive suffixes.

	Imp Ind Act	Imp Ind Pas	Imp Fut Act	Imp Fut Pas
1 sg	bam	bar	bo	bor
2 sg	bas	baris, bare	bis	beris, bere
3 sg	bat	batur	bit	bitur
1 pl	bamus	bamur	bimus	bimur
2 pl	batis	bamini	bitis	bimini
3 pl	bant	bantur	bunt	buntur

Table 4.34: Some infectum desinences of the verb 'amo' (to love) (repeated)

However, TAM markers in table 4.33, repeated here as table 4.35, do show a systematic syntagmatic relationship with affixes after them, since it is always the case that when 'era', 'isse' and 'eri' appear, they are followed by the same affixes.

	Pluperf Ind	Pluperf Sub	Perf Sub
1 sg	eram	issem	erim
2 sg	eras	isses	eris
3 sg	erat	isset	erit
1 pl	eramus	issemus	erimus
2 pl	eratis	issetis	eritis
3 pl	erant	issent	erint

Table 4.35: Some perfectum desinences of the verb 'amo' (to love) (repeated)

The proposal we make to include the effects of the MPC in Pinker (1984)'s model is the following:

(a) Once the stem is segmented, the procedure analyses syntagmatic relationships between the affixes in word-specific paradigms built on the basis of sentences in the target language.

(b) If affixes show a non systematic syntagmatic relationship, the mechanism proceeds normally as in Pinker (1984). The learner chooses a dimension from a multidimensional word-specific paradigm and a level of that dimension, examines all the cells of the paradigm specified for that level, and extracts the common phonetic material. Then, he enters that material in a cell of a unidimensional general paradigm corresponding to that level. This mechanism is repeated for the other levels of that dimension to complete the paradigm. Once this is done, the learner repeats the entire process for other dimensions in order to build general paradigms for them. In the case that the learner applies the former procedure and finds no phonetic material in common, he chooses two dimensions simultaneously, considering the combinations of the levels of the two dimensions in order to search for common material. The learner applies this procedure for all the dimensions of the paradigm. Once this is done, if no phonetic material in common exists, affixes are extracted and placed in a general paradigm resembling the specific one.

(c) If affixes do show a systematic syntagmatic relationship, the learner has to extract all the material in cells and create a multidimensional general paradigm. In practice, then, if affixes show a systematic syntagmatic relationship in a word-specific paradigm, the learner turns it into a general paradigm without changing anything.

We can illustrate the relevant cases continuing with the Latin morphological system. First, the learner would analyse their word-specific paradigms in 4.34 and 4.35 in order to determine the systematicity in the syntagmatic relationships between the affixes in cells. He would discover, comparing forms, that in paradigm 4.34, affixes do not show a systematic syntagmatic relationship, and then he would apply the procedure as in Pinker (1984)'s system, building the following general paradigms:

	Singular	Plural
1	-o, -m	-mus
2	-s	-tis
3	-t	-nt

Table 4.36: Personal desinences (I, active)

	Singular	Plural
1	-r	-mur
2	-ris, -re	-mini
3	-tur	-ntur

Table 4.37: Personal desinences (II, passive)

Imp	Imp Fut
ba	b/bis

Table 4.38: Some TAM markers

Given that TAM affixes in table 4.35 do show a systematic syntagmatic relationship with respect to personal desinences, the learner directly extracts all material in cells and creates a multidimensional paradigm, that is, the learner converts the word-specific paradigms in a general paradigm for all the dimensions involved:

	Pluperf Ind	Pluperf Sub	Perf Sub
1 sg	eram	issem	erim
2 sg	eras	isses	eris
3 sg	erat	isset	erit
1 pl	eramus	issemus	erimus
2 pl	eratis	issetis	eritis
3 pl	erant	issent	erint

Table 4.39: General paradigm of some perfectum desinences of the verb 'amo' (to love)

Thus when affixes do not show a systematic syntagmatic relationship, Pinker (1984)'s model works as usual. However, when affixes show a systematic syntagmatic relationship we propose that the learner applies the MPC and, given our concrete example, instead of postulating two paradigms, that is, one for the TAM markers *era, isse, eri* and one for personal desinences $\{m, ..., nt\}$, he prefers to postulate only one for all the information.

As is clear from the proposal, in order to introduce the effects of the MPC in the acquisitional model of morphological paradigms developed by Pinker (1984), it is necessary for the mechanism that inspects affixes left in word-specific paradigms to take into consideration the syntagmatic contexts where the forms appear. Actually the feature of inspecting if other morphs can appear in between of two morphs under consideration, as in the above mentioned case of *the big dog*, is also present in the original model to distinguish free and bound morphemes. Hence the actual implementation of the MPC makes use of learning mechanisms already present in the original system.

Recall that it has been pointed out that agglutinative patterns are more easily acquired than fusional patterns. Pinker (1984)'s original model accounts for these data by appealing to the big processing capacity needed by the learner when transferring a fusional word-specific paradigm to a general paradigm. Contrarily, the learner requires less processing capacity when doing the same with an agglutinative paradigm, as explained in the previous section. However, this argument cannot be adduced in our proposal: in the model we have proposed, what Pinker names fusional affixes are, in the proposed configuration, just extracted and placed in a general paradigm of the same form. In other words, given our proposal word-specific paradigms whose affixes show a systematic syntagmatic relationship turn into general paradigms without any modification. This operation does not seem *a priori* to be difficult for the learner to undertake. Consequently, we have to appeal to another reason if we want to account for the developmental data that show that agglutinative patterns are easier and more quickly acquired.

With respect to this, consider the following statement by Pinker regarding the creation of general paradigms from word-specific ones for agglutinative and for fusional affixes:

> For an agglutinative language, the one-dimensional general paradigm is strengthened several times in the course of one complete pass through the word-specific paradigm; for a fusional language it would only be strengthened once in the course of a complete pass (Pinker 1984, 191).

Pinker suggests, then, that the strength of a morph in a paradigm depends on the number of times this morph appears. Let us exemplify what this means with the abstract paradigms used above, repeated here for convenience:

	Nom	Acc
Sg	ip	iq
Pl	jp	jq

Table 4.40: Idealised agglutinative paradigm (repeated)

	Nom	Acc
Sg	a	b
Pl	c	d

Table 4.41: Idealised fusional Paradigm

The agglutinative suffixes in 4.40 are strengthened insofar as the procedure passes twice through each one: the singular 'i' is expressed in two cells, the nominative and the accusative cells, and the same holds for the plural 'j'. Likewise the nominative 'p' is expressed in two cells, the singular and the plural, and the same holds for the accusative 'q'. In this sense the resulting general paradigms, which we repeat in 4.42 and 4.43, are strengthened through the progression from specific to general paradigms:

Nom	Acc
p	q

Table 4.42: Agglutinative suffixes for case

Sg	Pl
i	j

Table 4.43: Agglutinative suffixes for number

On the contrary, when the procedure goes through the fusional paradigm in 4.41 none of the affixes appear more than once, and in this sense Pinker argues that the general resulting paradigm, repeated in 4.44, is not strengthened:

	Nom	Acc
Sg	a	b
Pl	c	d

Table 4.44: Fusional suffixes for case and number

We propose that this argument can also be used in our system to explain the strong difference between the developmental data on the acquisition of agglutinative and fusional morphology: given that agglutinative general paradigms are more strengthened than general fusional paradigms, their acquisition by the learner is easier and quicker (we will briefly come back to the argument of strengthened paradigms in section 4.3 when discussing the relationship between the present proposal and patterns of linguistic variation and change).

Let us focus on the proposed restriction, the MPC, and its role in the acquisition of inflectional systems in natural languages. As we have argued and as Pinker recognises, a mechanism of inspecting syntagmatic relationships between morphs must be at work during the process of language acquisition on independent grounds, since it is a mandatory task for the learner to acquire whether minimal meaningful pieces in his input language are free or bound (for more on the relevance of that distinction see chapter 5). The MPC is, thus, a restriction on how the acquisition of inflectional affixes proceeds that uses an element already present in the LAD. Moreover, the MPC works as a simplifying complexity device.

In order to develop this last point, we assume that a system is more complex than another one if it contains more elements, that is, more "occurrences of a variety of hand-picked, intuitively justified properties, or complexity indicators" (Bane 2008, 69). This intuitive definition of complexity has been used in several linguistic studies (McWhorter 2001, Roberts and Roussou 2003, Bickel and Nichols 2005, Shosted 2006) and this is also the sense which Roberts refers to when explaining that "the commonest way of determining this [complexity] is by simply counting some aspect of a derivation or

representation" (Roberts 2007, 234).[8]

For our purposes it is enough to take for granted this intuitive definition in order to argue how the MPC simplifies the complexity of the hypothesis carried out by the learner in the acquisition task. Let us specify two related ways in which the MPC simplifies the hypothesis with respect to Pinker (1984)'s original proposal.

First, given the effects of the MPC the learner hypothesises the smallest number of general paradigms, just one, in order to successfully analyse his input data when affixes show a concrete pattern. In Pinker (1984)'s original model it is proposed that the learner builds as many different general paradigms as needed by the system when acquiring morphological affixes, as we have exemplified with Latin verbal tenses. However, in this section we have proposed that the learner only has to build one general paradigm (containing TAM markers together with personal desinences) for all perfectum tenses because their affixes show the concrete pattern of being in a systematic syntagmatic distribution. Our learner has in the end a smaller number of general paradigms than the original learner, although he can anyway analyse his input data and acquire the inflectional system.

Second, our learner imposes a simpler representation to the input at the same time that he acquires the same underlying semantic distinctions. The original learner, as far as he builds one general paradigm for TAM markers and another for personal desinences, segments each verbal form in three pieces: *amav-isse-t*, for instance. However, our learner does not segment in three but in two the same verbal form: *amav-isset*, though he acquires all the underlying morphemes involved.

The MPC is a simplifying complexity device that enables the learner to carry out simple hypotheses to acquire his input language. Given that, the MPC can be considered by definition an element of the Third Factor

[8]Indeed this intuitive notion of complexity does not essentially differ from some formalised notions of this concept, as Kolmogorov complexity (Kolmogorov 1965). Kolmogorov complexity, which roughly states that the complexity of a string is the length of the shortest possible description of the string, is indeed based on this starting observation, though we are not going to go through them.

component in the growth of language in the individual that benefits efficient computation (Chomsky 2005).

4.3 The degree of fusion in verbal morphology

The constraint proposed in the previous section is plausible from an acquisitional perspective since it favours efficient computation and drives the learner towards simpler representations. However, we can also test the plausibility of the MPC beyond these theoretical arguments by considering empirical data. In this section I will make clear which predictions are derived from the effects of the MPC regarding morphological patterns in natural languages, and in the following section 4.4 I will show how these predictions are empirically confirmed. I will focus on verbal morphology, though most of these results may be extended to other domains.

There are two main predictions about the verbal morphology of natural languages derived from the effects of the MPC.

Given that we have proposed that the learner analyses his linguistic input differently according to whether or not he can apply the MPC, his resulting acquired language will be affected. Following this logic, we can argue that it is possible to observe some linguistic changes in natural languages derived from the effect of the successive analyses of learners depending on whether or not they apply the MPC when building the morphological paradigms of their target languages.

These predictions concern the degree of fusion in morphology (Sapir 1921). Sapir (1921) proposes two dependent indexes to classify languages of the world,[9] the degree of fusion and the degree of synthesis, although he recognises that the last one is not always useful. Let us focus on the degree of

[9]It is interesting to note that Sapir argues that the classification of languages "has much greater value if it is taken to refer to the expression of relational concepts alone" (Sapir 1921, 126), which may refer, in modern terminology, to the expression of functional/grammatical categories. This idea is not only current but it is closely related to the Borer-Chomsky Conjecture (BCC).

fusion, then. The degree of fusion refers to the "technique" or the "method" by which morphs relate to each other in a given language, that is, how they are assembled with respect to the others, and how they express grammatical morphemes. According to this criterion, it is possible to distinguish between an agglutinative pattern and an inflective or fusional pattern.[10]

The two proposed predictions about patterns of linguistic variation derived from the effects on the MPC when the learner is acquiring his target language are the following:

(9) **Prediction 1**: when some morphs show a systematic syntagmatic relationship in a paradigm and consequently the learner builds just one general paradigm for all the affixes, these affixes will (eventually) show a fusional pattern in subsequent instances of the language.

The logic behind prediction 1 is that, once affixes are put together in the same general paradigm, they will begin to show inconsistencies among forms, suffer morphophonological erosion and finally fusionalisation, due to the systematic contiguity of the pieces (all these cases are exemplified in the next section). For that reason, once a morph shows a systematic syntagmatic relationship with another morph, the two morphs will tend to undergo a process of fusionalisation. Therefore, in subsequent stages of the language, there will be just one fusionalised morph, due to successive analyses carried out by learners of different generations. This resulting morph will encode the same semantic distinctions previously encoded by the two morphs.

The weak degree of strength in fusional paradigms we mentioned in the previous section also favours the appearance of inconsistencies among forms and the morphophonological erosion of the affixes belonging to these paradigms.

(10) **Prediction 2**: when two morphs do no show such a systematic relationship and consequently the learner builds as many general

[10]Morphs can also show an isolating pattern if they are maximally independent of each other. However, I will not explore isolating patterns inasmuch as the MPC does not have effects on the acquisition of this type of morphs.

> paradigms as needed, the affixes will show an agglutinative pattern in subsequent instances of the language.

Prediction 2 captures the observation that discontiguity between morphs (and consequently the construction of more than one general paradigm by the learner) block morphophonological erosion and favours agglutination. Actually, this observation seems to be corroborated by the great conservativity shown by discontinuos patterns in Semitic languages.

These considerations suggest that the principles of analysis active in language acquisisition mould the I-language that will be finally acquired. If learners of a given generation introduce changes in the analyses of the target grammar consistently, then the language they will attain will be a modified version of the grammar underlying their linguistic input. In turn, this modified language will underlie the linguistic input received by the following generation of learners, and so on, giving as a result diachronic change. Indeed the dynamics of this process is basically the same as the one proposed to account for parametric change, as explained for example in Roberts (2007).

Before concluding this section, let us make clear what cases would contradict this model and, therefore, we expect not to find. First, we do not expect to find a scenario where two morphs show a systematic syntagmatic relationship and do not eventually become more inconsistent across the paradigm, eroded and fusionalised with each other. Second, we do not expect to find a context in which two morphs that do not show a systematic syntagmatic relationship are not agglutinative and also display extensive inconsistencies across the different forms of paradigm.

Sapir suggests that perhaps psychologists can discover why linguistic patterns are formed throughout different world languages and historical periods:

> Analogous [linguistic] trends are observable in remote quarters of the globe. From this it follows that broadly similar morphologies must have been reached by unrelated languages, independently and frequently (...) As linguists we shall be content to realize that there are these types and that certain processes in the life of language tend to modify them. Why similar types should be

> formed, just what is the nature of the forces that make them and dissolve them -these questions are more easily asked than answered. Perhaps the psychologists of the future will be able to give us the ultimate reasons for the formation of linguistic types (Sapir 1921, 121-122).

In this study we propose that we need to investigate learnability constraints and learning mechanisms that guide the process of language acquisition in order to find the causes of the origination of these types of linguistic variation and change.

4.4 Deriving patterns of linguistic variation

In this section I will present two case studies in order to illustrate the two predictions derived from the effects of the MPC: in section 4.4.1 I will show that the higher degree of fusionalisation of Romance verbal morphology, in particular of Catalan, Spanish, Italian and French, can be explained by appealing to prediction 1 above; in section 4.4.2 I will argue that the agglutinative nature of verbal pieces in Chichewa is a concrete example of prediction 2. Finally, in 4.4.3 I will present two other case studies considered for further research, regarding verbal morphology in Hua and in Turkish, to argue that they are good *prima facie* exponents of languages following prediction 1 and prediction 2 respectively.

4.4.1 Fusional patterns: from Latin to some Romance languages

As has been already mentioned, the goal of this part is, on the one hand, to show that verbal morphology of Catalan, Spanish, Italian and French show a higher degree of fusionalisation than Latin verbal morphology and, on the other, to argue that this is due to the effects of successive analyses made by learners using the MPC.

As Ledgeway (2012) explains, although Latin has traditionally been described as more synthetic than Romance, Romance continues to display extensive syntheticity involving such nominal and verbal inflections, as person, number, temporal, aspectual and modal categories on finite verbs.

I will present some verbal paradigms in Latin (4.4.1.1), and in Catalan (4.4.1.2), Spanish (4.4.1.3), Italian (4.4.1.4) and French (4.4.1.5). I have chosen the dialectal varieties more similar to the standard of each language, but it seems that the overall results obtained in this part can also be extended to the other dialects of the four languages.

I will show that regarding verbal affixes, a paradigmatic example of syntheticity in grammatical expressions, Romance shows more fusionalised paradigms than Latin. For simplicity, I will consider five simple verbal tenses, specifically, those that survive in Romance with the same underlying morphology as in Latin, namely, the present (pres.), the imperfect (imp.) and the perfect (perf.) indicative, and the present subjunctive (presS.); and one tense already present in Latin which took, in Romance varieties, the verbal TAM markers from other Latin tenses, that is, the imperfect subjunctive (impS.)(Alkire and Rosen 2010). I will be concerned with regular paradigms in all languages.

4.4.1.1 Latin

As has been already pointed out, verbal paradigms in Latin are maximally robust inasmuch as for all tenses and persons TAM markers and person/number desinences can be perfectly distinguished and isolated, and they have the same form across all paradigms. In all cases the verbal form is constituted by the stem, the TAM marker and the person/number affix.

In the following tables, Latin conjugations for the active voice are exemplified with the verbs *Amo* (stem *am-* and *amav-*), *Moneo* (stem *mone-* and *monu-*), *Rego* (stem *reg-* and *rex-*), *Capio* (stem *capi-* and *cep-*) and *Audio* (stem *audi-* and *audiv-*).

	Pres	Imp	Perf	PresS	ImpS
1 sg	Amo	Amabam	Amavi	Amem	Amarem
2 sg	Amas	Amabas	Amavisti	Ames	Amares
3 sg	Amat	Amabat	Amavit	Amet	Amaret
1 pl	Amamus	Amabamus	Amavimus	Amemus	Amaremus
2 pl	Amatis	Amabatis	Amavistis	Ametis	Amaretis
3 pl	Amant	Amabant	Amaverunt/ere	Ament	Amarent

Table 4.45: Latin verbal paradigm, 1st conjugation (*Amo, amas, amare, amavi, amatum*: 'to love')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Moneo	Monebam	Monui	Moneam	Monerem
2 sg	Mones	Monebas	Monuisti	Moneas	Moneres
3 sg	Monet	Monebat	Monuit	Moneat	Moneret
1 pl	Monemus	Monebamus	Monuimus	Moneamus	Moneremus
2 pl	Monetis	Monebatis	Monuistis	Moneatis	Moneretis
3 pl	Monent	Monebant	Monuerunt/ere	Moneant	Monerent

Table 4.46: Latin verbal paradigm, 2nd conjugation (*Moneo, mones, monere, monui, monitum*: 'to warn')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Rego	Regebam	Rexi	Regam	Regerem
2 sg	Regis	Regebas	Rexisti	Regas	Regeres
3 sg	Regit	Regebat	Rexit	Regat	Regeret
1 pl	Regimus	Regebamus	Reximus	Regamus	Regeremus
2 pl	Regitis	Regebatis	Rexistis	Regatis	Regeretis
3 pl	Regunt	Regebant	Rexerunt/ere	Regant	Regetent

Table 4.47: Latin verbal paradigm, 3rd (I) conjugation (*Rego, regis, regere, rexi, rectum*: 'to control')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Capio	Capiebam	Cepi	Capiam	Caperem
2 sg	Capis	Capiebas	Cepisti	Capias	Caperes
3 sg	Capit	Capiebat	Cepit	Capiat	Caperet
1 pl	Capimus	Capiebamus	Cepimus	Capiamus	Caperemus
2 pl	Capitis	Capiebatis	Cepistis	Capiatis	Caperetis
3 pl	Capiunt	Capiebant	Ceperunt/ere	Capiant	Caperent

Table 4.48: Latin verbal paradigm, 3rd (II) conjugation (*Capio, capis, capere, cepi, captum*: 'to take')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Audio	Audiebam	Audivi	Audiam	Audirem
2 sg	Audis	Audiebas	Audivisti	Audias	Audires
3 sg	Audit	Audiebat	Audivit	Audiat	Audiret
1 pl	Audimus	Audiebamus	Audivimus	Audiamus	Audiremus
2 pl	Auditis	Audiebatis	Audivistis	Audiatis	Audiretis
3 pl	Audiunt	Audiebant	Audiverunt/ere	Audiant	Audirent

Table 4.49: Latin verbal paradigm, 4th conjugation (*Audio, audis, audire, audivi, auditum*: 'to hear')

We repeat here for convenience TAM markers and person/number desinences in the active voice (the imperative excluded):

Pres	Imp	Perf	PresS	ImpS
⊘	ba	⊘	e (1st conj.)/a	re

Table 4.50: Som Latin TAM markers

	Singular	Plural
1	-o, -m	-mus
2	-s	-tis
3	-t	-nt

Table 4.51: Latin person and number desinences

	Singular	Plural
1	-i	-imus
2	-isti	-istis
3	-it	-erunt, -ere

Table 4.52: Latin person and number desinences for the perfect indicative

Consequently, as can be observed from the conjugation tables, Latin shows a clear agglutinative pattern in verbal affixes: they are completely robust and consistent, and, as a result, for each and every verbal form it is possible to isolate the TAM marker and the personal desinence, which are the same across all forms and persons.

Let us compare this picture with the verbal morphology displayed by the selected Romance languages: Catalan, Spanish, Italian and French. What we will see is that none of these Romance languages show verbal paradigms as robust as Latin verbal paradigms, though the degree of fusionalisation of affixes in these languages is in different stages. As will be observed, whereas Catalan personal desinences can still be distinguished from TAM markers although affixes are not consistent across all paradigms, in French the fusionalisation between TAM markers and personal desinences is almost total.

4.4.1.2 Catalan

In the following tables, Catalan conjugations are illustrated with the verbs *Treballar* (stem *treball-*), *Témer* (stem *tem-*) and *Dormir* (stem *dorm-*).[11]

	Pres	Imp	Perf	PresS	ImpS
1 sg	Treballo	Treballava	Treballí	Treballi	Treballés
2 sg	Treballes	Treballaves	Treballares	Treballis	Treballessis
3 sg	Treballa	Treballava	Treballà	Treballi	Treballés
1 pl	Treballem	Treballàvem	Treballàrem	Treballem	Treballéssim
2 pl	Treballeu	Treballàveu	Treballàreu	Treballeu	Treballéssiu
3 pl	Treballen	Treballaven	Treballaren	Treballin	Treballéssin

Table 4.53: Catalan verbal paradigm, 1st conjugation (*Treballar*: 'to work')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Temo	Temia	Temí	Temi	Temés
2 sg	Tems	Temies	Temeres	Temis	Temessis
3 sg	Tem	Temia	Temé	Temi	Temés
1 pl	Tememe	Temíem	Temérem	Tememe	Teméssim
2 pl	Temeu	Temíeu	Teméreu	Temeu	Teméssiu
3 pl	Temen	Temien	Temeren	Temin	Temessin

Table 4.54: Catalan verbal paradigm, 2nd conjugation (*Témer*: 'to be afraid of')

[11] In Catalan, the perfect indicative is not very productive in all forms, and its frequency of use is different across dialectal variants.

	Pres	Imp	Perf	PresS	ImpS
1 sg	Dormo	Dormia	Dormí	Dormi	Dormís
2 sg	Dorms	Dormies	Dormires	Dormis	Dormissis
3 sg	Dorm	Dormia	Dormí	Dormi	Dormís
1 pl	Dormim	Dormíem	Dormírem	Dormim	Dormíssim
2 pl	Dormiu	Dormíeu	Dormíreu	Dormiu	Dormíssiu
3 pl	Dormen	Dormien	Dormiren	Dormin	Dormissin

Table 4.55: Catalan verbal paradigm, 3rd conjugation (*Dormir*: 'to sleep')

In Catalan the person and number suffixes are not fused with the TAM markers, since it is possible to distinguish the following endings for each verbal form:

1 sg	⊘
2 sg	-s
3 sg	⊘
1 pl	-m
2 pl	-u
3 pl	-n

Table 4.56: Catalan desinences for person and number

Nevertheless, Catalan paradigms are not as robust as Latin ones, since TAM markers are more inconsistent and change from form to form (Mascaró 1986). For example, in the first and second conjugation the TAM marker for the present subjunctive is 'i' for some forms and 'e' for others: *treballi* ('I/(s)he work.subjunctive') and *temi* ('I/(s)he be-afraid-of.subjunctive'), but *treballem* ('We work.subjunctive') and *temeu* ('You.plural be afraid of.subjunctive').

Moreover, the perfect also shows diverse patterns of realisation, namely, 'í' , 'à', 'é' and 're': *treballí* ('I worked'), *treballà* ('(S)he worked'), *temè* ('(S)he was afraid of') and *dormires* ('You.singular slept').

If it is true that in Latin there are some alternation of TAM markers among conjugations and persons (for instance in the realisation of the present subjunctive with 'e' in the first conjugation and 'a' in the rest), this is not as extended as in Catalan. This fact is an indicator that the fusionalisation process is taking place: the paradigms stop being completely consistent and it is not possible to isolate TAM markers to be found in all forms.

4.4.1.3 Spanish

In the tables below Spanish morphological paradigms for each conjugation are exemplified with the verbs *Amar* (stem *am-*), *Temer* (stem *tem-*) and *Partir* (stem *part-*).

	Pres	Imp	Perf	PresS	ImpS
1 sg	Amo	Amaba	Amé	Ame	Amara
2 sg	Amas	Amabas	Amaste	Ames	Amaras
3 sg	Ama	Amaba	Amó	Ame	Amara
1 pl	Amamos	Amábamos	Amamos	Amemos	Amáramos
2 pl	Amáis	Amabais	Amasteis	Améis	Amarais
3 pl	Aman	Amaban	Amaron	Amen	Amaran

Table 4.57: Spanish verbal paradigm, 1st conjugation (*Amar*: 'to love')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Temo	Temía	Temí	Tema	Temiera
2 sg	Temes	Temías	Temiste	Temas	Temieras
3 sg	Teme	Temía	Temió	Tema	Temiera
1 pl	Tememos	Temíamos	Temimos	Temamos	Temiéramos
2 pl	Teméis	Temíais	Temisteis	Temáis	Temierais
3 pl	Temen	Temían	Temieron	Teman	Temieran

Table 4.58: Spanish verbal paradigm, 2nd conjugation (*Temer*: 'to be afraid of')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Parto	Partía	Partí	Parta	Partiera
2 sg	Partes	Partías	Partiste	Partas	Partieras
3 sg	Parte	Partía	Partió	Parta	Partiera
1 pl	Partimos	Partíamos	Partimos	Partamos	Partiéramos
2 pl	Partís	Partíais	Partisteis	Partáis	Partierais
3 pl	Parten	Partían	Partieron	Partan	Partieran

Table 4.59: Spanish verbal paradigm, 3rd conjugation (*Partir*: 'to leave/to divide')

As can be observed in the tables above, Spanish also shows more fusionalised verbal affixes than Latin. First, Spanish paradigms are more inconsistent in the TAM markers used across persons than Latin paradigms. Notice, for example, the TAM morphs used to mark the perfect, which include the exponents 'é', 'ste', 'ó' and 'í': *amé* ('I loved'), *amaste* ('You.singular loved'), *temió* ('(S)/he was afraid of') and *partí* ('I left').

Second, the desinence of the second person singular in the perfect tense is fusionalised with the TAM, in the sense that the etymological 's', still present in Catalan and in the other Spanish tenses, has disappeared: *amaste*, *temiste* and *partiste*.[12]

Thus in Spanish personal desinences can be distinguished from TAM markers with the exception of the second person, since *-s* is not present in all forms: in the perfect indicative is "fused" with the TAM. The resulting table of desinences in Spanish would be the following:

[12]However, there are some Spanish varieties where the 's' is pronounced: 'amastes', 'temistes' and 'partistes'. Actually this may be due to the strong power of analogy in building morphological paradigms and generally in acquisition (Blevins and Blevins 2009).

1 sg	⊘
2 sg	-s/⊘
3 sg	⊘
1 pl	-mos
2 pl	-is
3 pl	-n

Table 4.60: Spanish desinences for person and number

4.4.1.4 Italian

In the tables below Italian verbal paradigms are illustrated with the verbs *Amare* (stem *am-*), *Credere* (stem *cred-*) and *Dormire* (stem *dorm-*). The perfect or "passato remoto" is included since it is used in central and south variaties of spoken Italian (as well as in written standard Italian).

	Pres	Imp	Perf	PresS	ImpS
1 sg	Amo	Amavo	Amai	Ami	Amassi
2 sg	Ami	Amavi	Amasti	Ami	Amassi
3 sg	Ama	Amava	Amò	Ami	Amasse
1 pl	Amiamo	Amavamo	Amamo	Amiamo	Amassimo
2 pl	Amate	Amavate	Amaste	Amiate	Amaste
3 pl	Amano	Amavamo	Amarono	Amino	Amassero

Table 4.61: Italian verbal paradigm, 1st conjugation (*Amare*: 'to love')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Credo	Credevo	Credei	Creda	Credessi, Credetti
2 sg	Credi	Credevi	Credesti	Creda	Credessi
3 sg	Crede	Credeva	Credé	Creda	Credesse, Credette
1 pl	Crediamo	Credevamo	Credemmo	Crediamo	Credessimo
2 pl	Credete	Credevate	Credeste	Crediate	Credeste
3 pl	Credeno	Credevano	Crederono	Credano	Credessero, Credettero

Table 4.62: Italian verbal paradigm, 2nd conjugation (*Credere*: 'to believe')

	Pres	Imp	Perf	PresS	ImpS
1 sg	Dormo	Dormivo	Dormii	Dorma	Dormissi
2 sg	Dormi	Dormivi	Dormisti	Dorma	Dormissi
3 sg	Dorme	Dormiva	Dormí	Dorma	Dormisse
1 pl	Dormiamo	Dormivamo	Dormimo	Dormiamo	Dormissimo
2 pl	Dormite	Dormivate	Dormiste	Dormiate	Dormiste
3 pl	Dormono	Dormivano	Dormirono	Dormano	Dormissero

Table 4.63: Italian verbal paradigm, 3rd conjugation (*Dormire*: 'to sleep')

As can be observed in the data in the tables, in Italian TAM markers are also more inconsistent across the paradigm than Latin ones, and in some cases, it is even difficult to take them apart from personal desinences, specially in the 1sg, 2sg and 3sg of all tenses and conjugations.

The higher degree of fusionalisation in verbal suffixes is easily observable, for example, in the 2sg of the present subjunctive in the second and third conjugations, which are *creda* and *dorma*: this last 'a' does not appear in any other form of the paradigms associated with the 2sg.

Also the personal desinence for the 3pl show four different exponents, one with the imperfect 'mo' *amavamo* ('They used to love'), one with the perfect 'rono' *amarono* ('They loved'), one with the imperfect subjunctive 'ero' *amassero* ('They loved.subjunctive') and another one with the two presents 'no' *amano, amino* ('They love, they love.subjunctive').

It would be very difficult to build the complete paradigm of the personal desinences given the inconsistency of the system, though the following regularities can be extracted:

1 pl	-mo
2 pl	-te

Table 4.64: Some Italian desinences for person and number

4.4.1.5 French

In the following tables there are French verbal paradigms, which only include the present, the imperfect and the present subjunctive, given that the other two simple tenses considered before are generally not used in spoken French (they are usually replaced by analytic constructs). The verbs are *Chanter* (stem *chant-*), *Mettre* (stem *me/ɛt-*) and *Finir* (stem *fin-/finis-*). Both orthographic and phonetic paradigms are provided, due to the strong disparity in French between them.

	Pres	Imp	PresS
1 sg	Chant⊘	Chantɛ́	Chant⊘
2 sg	Chant⊘	Chantɛ́	Chant⊘
3 sg	Chant⊘	Chantɛ́	Chant⊘
1 pl	Chantɔ̃	Chantjɔ̃	Chantjɔ̃
2 pl	Chante	Chantjé	Chantjɛ́
3 pl	Chant⊘	Chantɛ́	Chant⊘

Table 4.65: French verbal paradigm (phonetic), 1st conjugation (*Chanter*: 'to sing')

	Pres	Imp	PresS
1 sg	Mɛ́	Metɛ́	Mɛ́
2 sg	Mɛ́	Metɛ́	Mɛ́
3 sg	Mɛ́	Metɛ́	Mɛ́
1 pl	Metɔ̃	Metjɔ̃	Metjɔ̃
2 pl	Meté	Metjɛ́	Metjɛ́
3 pl	Mɛ́t	Metɛ́	Mɛ́t

Table 4.66: French verbal paradigm (phonetic), 2nd conjugation (*Mettre*: 'to put')

	Pres	Imp	PresS
1 sg	Fini	Finisɛ́	Finis∅
2 sg	Fini	Finisɛ́	Finis∅
3 sg	Fini	Finisɛ́	Finis∅
1 pl	Finisɔ̃	Finisjɔ̃	Finisjɔ̃
2 pl	Finisɛ́	Finisjɛ́	Finisjɛ́
3 pl	Finis∅	Finisɛ́	Finis∅

Table 4.67: French verbal paradigm (phonetic), 3rd conjugation (*Finir*: 'to end')

	Pres	Imp	PresS
1 sg	Chante	Chantais	Chante
2 sg	Chantes	Chantais	Chantes
3 sg	Chante	Chantait	Chante
1 pl	Chantons	Chantions	Chantions
2 pl	Chantez	Chantiez	Chantiez
3 pl	Chantent	Chantaient	Chantent

Table 4.68: French verbal paradigm (orthographic), 1st conjugation (*Chanter*: 'to sing')

	Pres	Imp	PresS
1 sg	Mets	Mettais	Mette
2 sg	Mets	Mettais	Mettes
3 sg	Met	Mettait	Mette
1 pl	Mettons	Mettions	Mettions
2 pl	Mettez	Mettiez	Mettiez
3 pl	Mettent	Mettaient	Mettent

Table 4.69: French verbal paradigm (orthographic), 2nd conjugation (*Mettre*: 'to put')

	Pres	Imp	PresS
1 sg	Finis	Finissais	Finisse
2 sg	Finis	Finissais	Finisses
3 sg	Finit	Finissait	Finisse
1 pl	Finissons	Finissions	Finissions
2 pl	Finissez	Finissiez	Finissiez
3 pl	Finissent	Finissaient	Finissent

Table 4.70: French verbal paradigm (orthographic), 3rd conjugation (*Finir*: 'to end')

As can be observed in the phonetic paradigms, French personal desinences and TAM markers show a very high degree of fusion, and even in some cases verbal suffixes have almost disappeared. As a consequence, it is not possible to draw a table just for desinences. In this respect, the difference between French affixal system and Latin is very remarkable.

Having showing in which sense Romance verbal paradigms show a higher degree of fusionalisation than Latin ones, let us focus on how is it possible to explain this fusionalisation appealing to the successive analyses of learners using the MPC.[13]

[13]What it is argued for here is that the effects of the MPC on affixes that show a

The layout of the argument is the following: (a) verbal affixes in Latin, namely, TAM markers and personal desinences, did not show a systematic syntagmatic relationship due to the existence of two sets of personal desinences, one for the active voice and one for the passive voice, (b) Latin suffered a change in its passive system and began to construct all passive sentences analytically, without resorting to the passive set of personal desinences, (c) this change implied that the TAM markers and the personal (active) desinences started to show a systematic syntagmatic relationship, (d) from that point onwards learners could apply the MPC in their successive analyses, finally (e) the effects of the MPC gave as a result a more fusionalised system of verbal morphology in successive instances of Latin, that is, in Romance languages.

Let us illustrate these processes in what follows. The argumentation uses data already presented in the previous sections, that we repeat here for convenience.

(a) Infectum tenses in Latin (present indicative and subjunctive, imperfect indicative and subjunctive, and imperfect future) are formed by combining TAM markers with two sets of personal desinences, one for the active voice and another for the passive voice, as in the following tenses of the first conjugation:

systematic syntagmatic relationship can explain why verbal affixes in Romance are more inconsistent than in Latin and have the possibility of progressively losing phonetic material. It is important to make clear that this does not deny that the loss of verbal agreement may be connected with other linguistic phenomena in the system, as for instance the behaviour of subjects (Adams 1987, Roberts 1993, Kato 1999).

	Imp Ind Act	Imp Ind Pas	Imp Fut Act	Imp Fut Pas
1 sg	Amabam	Amabar	Amabo	Amabor
2 sg	Amabas	Amabaris, Amabare	Amabis	Amaberis, Amabere
3 sg	Amabat	Amabatur	Amabit	Amabitur
1 pl	Amabamus	Amabamur	Amabimus	Amabimur
2 pl	Amabatis	Amabamini	Amabitis	Amabimini
3 pl	Amabant	Amabantur	Amabunt	Amabuntur

Table 4.71: Some infectum forms of the verb 'amo' (to love)

The TAM markers in table 4.71, 'ba' and 'b'/'bis', do not show a systematic syntagmatic relationship with the affixes after them because it is not the case that these TAM markers are always followed by the same set of affixes: in the active paradigms TAM markers are followed by some affixes $\{m, ..., nt\}$, and in the passive paradigms TAM markers are followed by some others $\{r, ..., ntur\}$.

This situation is different in perfectum tenses, whose TAM markers do show a systematic syntagmatic relationship with personal desinences, due to the fact that the passive was formed analytically, as illustrated in example (8), repeated below.

(11) Equus vulneratus est hasta
horse.NOMsg injure.pp to-be.3sg spear.ABLsg
'The horse was injured by a spear.'
(Valentí i Fiol 1995, 106)

We will come back to perfectum tenses at the end of this section.

(b) In the Vulgar Latin of the late period, from the II AD ownwards, synthetic passive was gradually replaced by analytical passive in all tenses (Väänänen 1982, 208).

(c) Due to the analytical construction of the passive, all TAM markers, also infectum TAM affixes, began to show a systematic syntagmatic relationship with personal desinences, which are reduced to only one set for the active $\{m, ..., nt\}$.

(d) From the moment where TAM markers and personal desinences displayed a systematic syntagmatic distribution, the learner can apply the MPC to all tenses in their input language. In doing that when building morphological paradigms, the learner directly extracts all material in cells and creates a multidimensional paradigm, that is, the learner converts word-specific paradigms in general paradigms for all the dimensions involved. In other words, instead of postulating two paradigms, that is, one for the TAM markers and one for personal desinences, the learner postulates only one for all the information.

(e) This operation, which results in all the information being in the same paradigm and not in two different paradigms, not only favours the inconsistency of affixes across tenses and forms but also the phonological erosion of pieces, which produces the fusionalisation of affixes in subsequent instances of Latin, that is, in Romance languages.

According to the rationale we are exploring here, it would be expected to find fusionalised or contracted forms attested in Vulgar Latin in perfectum tenses, due to the effects of the MPC. Given that perfectum TAM markers always appeared followed by active desinences, since their passive was always analytically formed, the fusionalisation had to eventually take place in these tenses given the developed proposal.

As Herman (2000) explains, the attested differences between Classical Latin and Vulgar Latin regarding verbal morphology are scarce. Nonetheless, it seems symptomatic that the few contracted forms attested in Vulgar Latin do belong to perfectum forms, particularly those ending in -are. For instance, the following Vulgar Latin forms has been attested: *probai*, *probait*, *probaut* and *probat*. Their counterparts in Classical Latin are *probavi* ('I tried') for the contracted form *probai*, and *probavit* ('he tried') for the other three contracted

forms *probait*, *probaut* and *probat*. All these attested contracted forms are considered "spoken usages" (Herman 2000, 80).[14] These "spoken usages" may well be the result of the process we argue for in this section: since those TAM markers belonging to perfectum tenses did show a systematic syntagmatic relationship with personal desinences, learners could apply the MPC and its effects caused the fusionalisation of verbal affixes.

4.4.2 Agglutinative patterns: Chichewa and Bantu languages

The goal of this section is to argue that the agglutinative nature of verbal pieces in Chichewa, and generally in Bantu languages, can be derived from the successive analyses carried out by learners in the process of language acquisition.

We extensively draw on Mchombo (2001) for data and explanations on Chichewa. Chichewa is a Bantu language spoken in Malawy, where it is the national language, Mozambique, Zambia and Zimbabwe, and it is also known as "Chinyanja". Chichewa is a typical Bantu language in its morphological structure: it has an elaborate system of noun classification and shows highly agglutinative and complex verbal morphology.

For a verbal form to be well-formed in Chichewa, it must include in a simple sentence the following components: a) the verb root, b) the final vowel (fv), which is 'a', for the indicative mood, and 'e', for the subjunctive mood, c) the subject marker, and d) the tense/aspect marker. In some cases, the object marker is also mandatory.

In example (12), the verbal stem for 'smash', *phwány*, is just accompanied by the mandatory pieces: the agreed subject marker, the tense/aspect marker, the agreed object marker and the final vowel:

[14]It is important not to confuse these contracted forms in Vulgar Latin with the contracted forms in Classical Latin in perfectum tenses when -*vi*- preceds -*s*- and when -*ve*- precedes -*r*-, as in: *amavisti* ≻ *amasti*, *amaverunt* ≻ *amarunt* or *amaveras* ≻ *amaras*. These alternations are due to phonological factors (Valentí i Fiol 1995, 92).

(12) chi-ku-wá-phwány-a
7Smasc-PRES-6Omasc-smash-fv
'He (it) is smashing them.'
(Mchombo 2001, 501)

In example (13) both the subject, *chigawênga*, and the object, *maûngu*, are present in the sentence; apart from this, the stem is accompanied by the subject marker, the tense/aspect morph and the final vowel:

(13) chigawênga chi-ku-phwány-á maûngu
7.terrorist 7Smasc-PRES-smash-fv 6.pumpkins
'The terrorist is smashing some pumpkins.'
(Mchombo 2001, 500)

What is particularly interesting for us about Chichewa's system is that, given the full structure that the verbal form can display, none of the mandatory morphs of the verbal complex are always adjacent, that is, none of the mandatory morphs show a systematic syntagmatic relationship with another morph. The full structure of the verbal complex in simple sentences is the following (mandatory pieces and the verbal root are in bold):

(14) (Negation in simple clauses) + **Subject Marker** + (Negation in subordinate clauses) + **Tense/Aspect marker** + (prefixes as Modals, Directionals or Conditionals markers) + (Object Marker/Reflexive particle) + **Verbal root** + (suffixes or "extensions" such as the Causative, the Applicative, the Passsive, the Stative, or the Recriprocals marker) + **final vowel**.

In the following example, the verbal root 'smash', *phwány*, is preceded by the marker of the negation in subordinate clauses, *sa*. In cases of sentential complementation like (15) the final vowel *e* replaces the tense/aspect marker, which does not appear:

(15) Kalúlú a-ku-fún-á kutí mlângo u-sa-phwány-é
1.hare 1Smasc-PRES-want-fv that 3.lion 3Smasc-neg-smash-fv
maûngu
6.pumpkins

'The hare wants the lion not smash the pumpkins.'
(Mchombo 2001, 503)

This example suffices to show that the mandatory morph for the subject marker does not show a systematic syntagmatic relationship with the next mandatory morph in the structure: first, the negation marker appears following the subject marker in subordinate clauses; second, the tense/aspect marker, which is mandatory in simple sentences, does not appear in all clauses, as for example in sentential subordination.

In the following example in (16), the directional *ká* appears following the tense/aspect marker:

(16) mkângo u-ku-ká-phwány-a máûngu
3.lion 3Smasc-PRES-go-smash-fv 6.pumpkins
'The lion is going to smash some pumpkins.'
(Mchombo 2001, 503)

This example shows that the mandatory tense/aspect morph does not show a systematic syntagmatic relationship with the object marker, which is not even present in the example due to the fact that the full object 'pumpkins', *máûngu*, appears in the sentence.

In the example in (17) the passive maker *ídw* appears before the final vowel:

(17) maûbgu a-ku-phík-ídw-a
6.pumpkins 6Smasc-PRES-cook-pass.-fv
'The pumpkins are being cooked.'
(Mchombo 2001, 508)

The example above shows that the mandatory morph for the final vowel can be preceded by what is known as "extensions" in Bantu linguistics. Extensions are markers that affect the number and nature of arguments in the verbal configuration, such as the causative, the applicative or the passive.

What has been shown by the Chichewan examples is that none of the mandatory morphs of the verbal complex show a systematic syntagmatic distribution with respect to another morph. Given the proposed model for

the acquisition of verbal affixes, the learner of this system has to build as many general paradigms as needed, not just only one for all the adjacent information. Consequently, according to Prediction 2, the acquired affixes display an agglutinative pattern and will display it in subsequent instances of the language as long as morphs continue showing that specific configuration.

It is interesting to note that the argument developed for Chichewa can be extended to languages belonging to the Bantu family. As Güldemann (2003) argues, the traditional schema of the full structure of the verbal form in Bantu languages, as developed by Meeussen (1967), can be simplified as in (18):

(18) (preinitial) initial (postinitial) (preradical) radical (prefinal) final (postfinal)

As is clear from the slots available in the Bantu verbal form, none of the mandatory pieces is always adjacent, in other words, none of them shows a configuration of syntagmatic systematicity. Given that and assuming the model developed in this chapter, the agglutinative nature of Bantu verbal morphs can naturally be explained.

4.4.3 Other case studies

In this part of the chapter we are going to point out two cases that at first sight seem good candidates for further research on the effects of the MPC. Hua, a Gorokan language, seems to have suffered a process of fusionalisation that, though being very different from the one suffered by the Romance languages, may be explained by appealing to the effects of the MPC. Turkish shows a configuration of agglutinative verbal affixes that also seems to be the consequence of the dynamics of the MPC. Our objective here is to present the basic data of the verbal morphology in these languages in order to prove that they are good candidates for further research on the MPC and the two predictions derived from it. The scarce data available on Hua prevent me from considering the patterns of this language with the same detail as

the patterns considered previously. Admittedly, the rather complex Turkish morphological verbal system requires more scrutiny.

4.4.3.1 A Gorokan Language: Hua

This section is based on the studies of Hua by John Haiman (Haiman 1980, 2001). Hua is a Papuan Language of the Gorokan family, spoken by 3.000 people in the Eastern Highlands of Papua New Guinea. Hua and languages belonging to the Gorokan family (as Move, Yate, Kamano and more) are some of the best known and most studied non-Austronesian languages of New Guinea. As Haiman explains, Gorokan languages are SOV, predominantly suffixing and predominantly agglutinative, they lack grammatical gender and are very regular in their marking of nominal case and verbal tense, mood, aspect and subject-verb agreement.

What is of interest for our purposes about this linguistic family is that Gorokan languages "differ in the degree to which agglutinative expression of subject-verb agreement (invariable stem + personal desinence) has been replaced by a synthesis of these two morphemes" (Haiman 2001, 539).

For instance, in Gimi, another Gorokan language, the system is like the one attested in previous stages of Hua: in order to mark subject-verb agreement, the invariable verbal stem is combined with a personal desinence that agrees with the subject of the sentence. Nevertheless, currently in Hua this system has been replaced by another one where the verbal stem is already inflected agreeing with the subject and at the same time that is combined with a personal desinence also sensitive to some properties of the subject. Specifically, in Hua the person and number of the subject of the verb are marked through a combination of vocalic ablaut affecting the final segment of the verb stem plus an only partially specified personal desinence, which also indicates illocutive force.

In table 4.72 verbal conjugations in Hua, or types, are specified:

	Type I (to do)	Type II (to eat)	Type III (to give)
1 sg., 1 dl., 1 pl.	hu	do	mu
3 sg.	hi	de	mi
2 sg., 2/3 dl., 2/3 pl.	ha	da	mi

Table 4.72: Hua stems paradigm

As can be observed, verbal stems are sensitive to the number and person of the subject: there is a form for the first person (singular, dual and plural), a form for the third person singular, and another one for the rest, that is, for the second person singular, dual, and plural, and the third person dual and plural.

Verbal forms must then be combined with personal desinences, also sensitive to some subject properties:

	Assertive	Interrogative	Exclamatory
Unmarked	e	ve	mane
Dual	'e	've	'mane
Other	ne	pe	pane

Table 4.73: Hua interrogative desinences

The unmarked desinences are combined with subjects in the first person singular, the third person singular and the second and third person plural; the dual desinences are used with dual subjects; whereas the other desinence is combined with second person singular and the first person plural subjects.

Hence to create an inflected form in Hua what is minimally needed is the combination of a verbal stem and a personal desinence, as in *hue* 'I do', *dave* 'Do they eat?' or *mipane* 'You give!'.

As Haiman explains, "Comparative evidence suggests that stem ablaut arose through vowel crasis, or gunah, the relative unstable vowel chain sequence /...Vowel$_1$ + Vowel$_2$.../ (still comparatively well preserved in Gimi),

being replaced by /...Vowel$_3$+.../. Nevertheless, it is impossible to posit phonologically plausible abstract underlying representations of the first type in a synchronic grammar of Hua" (Haiman 2001, 539).

Unfortunately, I do not have more data on Hua and on Gimi in order to try to elucidate what happened in Hua and why verbal stems got fusionalised and encoded part of the information carried solely by personal desinences in previous stages of the language. What we may conjecture is that some factor in Hua (and not in Gimi) favours the fusionalisation of two pieces that appeared adjacently, such as perhaps the availability of the ablaut process in the language.

Related to that conjecture, one could come up with the question of why it is not the case that the stem and the personal desinence get fusionalised in the present stage of the Hua language. The answer to this questions is indeed related to the proposal explored in this chapter: since these two morphs do not show a systematic syntagmatic relationship, as will be exemplified, they cannot undertake any fusionalisation. In Hua the verbal stem can be followed by an auxiliar, in the example *ba*, which appears before the personal desinence *e*, as in (19):

(19) hu -ba -e
do.1sg -progressive.1sg -desinence.affirmative.unmarked
'I am doing' ('I was doing').
(Haiman 1980, 137)

The aorist form, that is, the form composed solely by the verbal stem and the personal desinence, is used to indicate past or present actions and states; however, to indicate some other aspect and/or tense, such as progressive, future or habitual, an auxiliary must appear between the verb stem and the desinence. In example (19), the progressive marker is inserted between them.

From the data presented in Hua it is clear that the partial fusionalisation that occurred in this language between the verbal stem and the former personal desinence is very different from the fusionalisation presented in other sections of this chapter, where the stem was not involved and the fusionalisation occurs between grammatical morphs. Actually, the proposed MPC

is relative to the acquisition of inflectional systems. Nevertheless it may be worth taking into consideration the case of Hua for two reasons. First, the comparison between Hua and Gimi is particularly interesting because they are historically related, but only Hua underwent fusionalisation. It would be relevant to investigate if some change occurred in Hua, but not in Gimi, or viceversa, that favours the fusionalisation in one system but not in the other (remember that in the case of Latin, the change in the expression of the passive voice was what triggered the systematic syntagmatic distribution of verbal affixes). Second, the data that have been presented support our claims regarding Prediction 2, which states that pieces that do not show a systematic syntagmatic relationship show an agglutinative pattern. In Hua the stem and the personal desinence do not appear fusionalised arguably because sometimes another marker can appear between the two pieces. Thus, the data in Hua show that the non-syntagmatic distribution of morphs plays a role even in a context sensitive to fusionalisation, which has already taken place between the stem and a former personal desinence.

4.4.3.2 A Turkic Language: Turkish

Turkish is a Turkic language with more than 60 million native speakers, mainly in Turkey. Turkish displays a high degree of agglutination, vowel harmony and is a SOV language. In this subsection we are going to provide some arguments to argue that the agglutinative nature of verbal pieces in Turkish is derived from the effects of the MPC.

As Sezer (2001) explains, in Turkish the verbal root can appear with many optional derivative suffixes, though not all combinations are possible; this creates the verb stem:

(20) Verb stem = Verb root + (causative type_1) + (causative type_2) + (passive) + (abilitative) + (negation) + (possibilitative) + (aorist) + (person)

Leaving aside these optional suffixes, a finite verb stem in Turkish must minimally contain a main tense marker and a personal agreement desinence.

Sezer (2001) proposes that tense markers in Turkish can be classified in three blocks, which can be called Tense-1, Tense-2 and Tense-3. The author argues that, assuming this classification of tense markers, the verbal form in Turkish displays the following configuration:

(21) Verb stem + Tense-1 + (Tense-2) + (Tense-3) + Personal agreement

That means that for a Turkish verbal form to be well formed it must contain a marker belonging to the set of Tense-1 and the personal agreement desinence, Tense-2 and Tense-3 markers being optional.

As Sezer (2001) specifies, Tense-1 forms include the definite witnessed past ('DI'), the subjunctive conditional ('sE'), the inferential past/present perfect ('mIs'), the continuous ('Iyor'), the future ('yEcEG'), the aorist ('Ir/Er'), the optional/subjunctive ('yE'), the necessitative ('mEli') and the continuous ('mEkte'); Tense-2 forms include the definite witnessed past ('i-DI/(y)DI'), the indicative conditional ('i-sE/(y)sE') and the inferential ('i-mIs/(y)mIs'); finally, Tense-3 forms are the indicative conditional ('i-sE/(y)sE') and the inferential ('i-mIs/(y)mIs') (Sezer 2001, 4).

In example (22) the verbal root *gel* 'to arrive' is followed by a Tense-1 marker and the personal desinence; in example (23) the verbal root *yap* 'to do' is followed by a Tense-1 marker, a Tense-2 marker and the personal desinence; finally, in example (24) the verbal root *gid* 'to go' is followed by a Tense-1, a Tense-2 and a Tense-3 markers, plus the personal agreement desinence:

(22) yeni gel -di -m
just arrive -PAST(T-1) -1sg
'I have just arrived.'
(Sezer 2001, 10)

(23) yap -mis -ti -m
do -infer.PAST(T-1) -PAST(T-2) -1sg
'I had done it.'
(Sezer 2001, 12)

(24) gid -ecek -mis -se -m
go -FUT(T-1) -infer.PAST(T-2) -ind.COND(T-3) -1sg.
'(If it is the case that they say) I will/would go.'
(Sezer 2001, 3)

Sezer (2001) proposes that the full semantics of the inflectional complex may be read off compositionally from the individual meanings of the tense affixes.

The morphological complexity of Turkish verbal form, however, goes beyond that: there are also three different sets of personal agreement desinences that are chosen depending on the tense marker(s) present in the concrete verbal form. In the table below the three different personal agreement paradigms for the indicative and the subjunctive moods are specified:

	Type_1	Type_2	Type_3
1 sg	-Im	-m	-(y)AyIm
2 sg	-sIn	-n	-(y)AsIn
3 sg	⊘	⊘	-(y)A(-sIn)
1 pl	-Iz	-k	-(y)AlIm
2 pl	-sInIz	-nIz	-(y)AsInIz
3 pl	-lAr	-lAr	-(y)Alar (-sInlAr)

Table 4.74: Turkish agreement paradigms
(Kornfilt 1997, 382)

The personal desinences belonging to the set of Type_1 constitute the paradigm with the widest distribution, that is, personal desinences in Type_1 are used in most simple tense forms, in non-verbal predicates, etc. Agreement markers of the Type_2 are only used with the past tense marker *-di* belonging to the group Tense-1 and with the conditional subjunctive *-sE* belonging to the group of Tense-2. Finally, personal desinences of the Type_3 are exclusively chosen when the optative subjunctive tense *-yE*, belonging to the Tense-1 group, is used.

Given this distribution of tense markers and personal agreement desinences, there is no morph in Turkish verbal form that shows a systematic syntagmatic relationship with another morph. If it is true that there are personal desinences dedicated to Tense-1 markers, as Type_2 (25) and Type_3 personal agreement paradigms (26), and there are also personal desinences dedicated to Tense-2 markers, as Type_2 paradigm (27), recall that Tense-1 markers can be followed by Tense-2 markers (28), and Tense-2 markers can be followed by Tense-3 markers (29); thus, the systematic syntagmatic distribution is avoided:

(25) Tense-1 + Type_2

(26) Tense-1 + Type_3

(27) Tense-2 + Type_2

(28) Tense-1 + Tense-2

(29) Tense-2 + Tense-3

Therefore, none of the potentially adjacent morphs shows a systematic syntagmatic relationship and, as predicted by the model developed here, Turkish verbal pieces are agglutinative.

4.5 Conclusions

The proposal developed in this chapter is a concrete instantiation of how Plato's problem and Greenberg's problem can be related in a substantive way: it has been argued that a constraint active during language acquisition is responsible for shaping a concrete aspect of the languages of the world.

To articulate the proposal, Pinker (1984)'s model on the acquisition of inflectional systems has been assumed: adult morphological knowledge is organised in paradigms and the task of the learner is to discover the paradigms of his input language.

We have added some modifications to the original model in order to introduce in to the system the effects of the Minimise Paradigms Constraint (MPC), a proposed learning restriction active during language acquisition that regulates how the learner builds up morphological paradigm representations. This constraint is postulated on independent grounds insofar as it is coherent with the argument extensively developed in the literature concerning how the learner proceeds: trying to hypothesise the simplest structural representations.

Our next step has been to relate the effects generated by the MPC in the resulting acquired languages and the degree of synthesis in verbal morphology (Sapir 1921). Specifically, we have proposed that when two morphs show a systematic syntagmatic relationship they will eventually show a fusional pattern, that is, an unstable configuration across tenses and persons and an erosion of phonetic material. However, when two morphs do not show a systematic syntagmatic relationship they will show an agglutinative pattern. The former option has been exemplified with Latin and Romance affixal systems; the latter, with Chichewa verbal pieces.

Two other cases have also been suggested for consideration and further research, the verbal systems in Hua and Turkish. I considered convenient to include these two cases into the general discussion, although these data require more scrutiny: Hua presents a fusionalisation very different from the fusionalisation studied with Latin and Romance verbal systems, and Turkish verbal system is very intricate and complex.

Before concluding, I want to point out three natural consequences of the proposal developed in this chapter.

First, the fact that we have specified a context when the learner is going to analyse his linguistic input in a specific fashion, namely, the context of syntagmatic systematicity between morphs, can be considered a step forward in predicting how and when morphological change is eventually going to happen; and when by hypothesis it cannot occur. This is a powerful and formalised tool in examining and predicting morphological change. Therefore, this model is able to predict the context in which some concrete changes are expected to happen, and the context in which a concrete change is banned.

This contradicts the following view, which is sometimes found in the literature regarding morphological change:

> It may well be that for morphological change, a general theory -that is, a predictive theory- is not even possible, and that all that can be done is to catalogue tendencies, which, however valid they may be, do not in any sense constitute inviolable predictions about what types of changes will necessarily occur in a given situation (Joseph 2001, 366).

The second consequence has to do with the direction of change or drift. It is clear that this proposal implies that there is a learning bias towards simpler representations, which means that there will be, as a result, a drift towards certain patterns of linguistic variation. This view is not only defended by many authors, such as Roberts (2007), but is also a logical consequence of theories assuming parametric schemata. Parametric schemata are understood in the literature as learning paths and typological biases: the learner is supposed to be conservative and only move down the schema if forced to by PLD, which means that, in case of ambiguity, he would not take the next step and would remain in the higher node of the schema. A theory like that implies that linguistics is in charge not only of describing how possible languages are, but also what features are more probable than others (against authors such as Newmeyer (2005)). In a model assuming parametric schemata, points of acquisition in the top of parametric hierarchies would correspond to widespread linguistic patterns or, in other words, to most probable patterns (those chosen by the learner in ambiguity cases).

Let us clearly state the predictions of the model presented in this chapter, where parametric schemata have not been assumed, regarding the drift. As mentioned before, in our proposal we do argue for a drift towards simple representations. However, it is not claimed that all languages will change towards the fusionalisation of their morphology. Instead, we have proposed a very concrete scenario where the fusionalisation is going to be expected: when two morphs show a systematic syntagmatic distribution. What we claim, then, is that a concrete pattern is going to change towards a more

fusionalised one only when it is found in a concrete configuration. Given that it is not possible to predict when two morphs are going to display the relevant configuration, the theory does not imply that fusionalisation patterns are cross-linguistically more probable than others morphological patterns. Instead, what the theory predicts is that fusionalisation patterns are more probable, indeed they are the only expected pattern, when two morphs show the concrete configuration of being in a systematic syntagmatic relationship.

Third, the present proposal leads us to the appealing conclusion that there are distributional hints that guide the learner in finding out whether a particular morph is synthetic or non-synthetic (for more on that distinction see chapter 5). Agglutinative patterns are defined as those where each morph realises a single semantic unit, or morpheme, and each morpheme is realised by a single morph. Agglutinative morphs are, then, non-synthetic. On the contrary, in a fusional pattern it is not possible by definition to isolate a morph to correspond to every semantic unit, because a morph conveys more than one morpheme. Thus fusional morphs can be defined as synthetic. Following our model, if two morphs show a non systematic syntagmatic relationship, the learner is going to build as many general paradigms as needed, which means that morphs will show an agglutinative pattern. However, if two morphs show a systematic syntagmatic relationship, the learner builds just one general paradigm for all affixes, a situation that will favour their fusionalisation in subsequent instances of the language.

According to this reasoning, it is desirable to argue that the learner is able to use this distributional information, that is, the syntagmatic configuration where morphs appear, in order to guess and acquire how many morphemes are encoded in the relevant morphs. In other words, our system predicts that there are distributional properties that the learner can exploit in order to discover the semantic content of morphs, one of the formidable research problems acquisition is faced with, as Pinker states it, and generally one of the most striking problems of natural language learning. As far as I am aware, this is the first attempt in the literature to relate the use of distributional properties, which are extensively used by language learners (Redington et al.

1998, Redington and Chater 1998), with the discovery of concrete semantic aspects of morphs, namely, their syntheticity or non-syntheticity.

It has been proposed in the literature, mainly by Harris (1968, 1970) and colleagues, what is known as the *distributional hypothesis*. As Sahlgren (2008) explains, this hypothesis exploits the correlations between distribution similarity and semantic similarity of linguistic entities, where semantic similarity makes reference to notions such as synonymy, antonymy or hyponymy. Given this hypothesis, in exploring the distribution of words, the learner can discover these types of relationships among units. Although that particular framework exploits distributional properties with the aim of discovering some semantic aspects of pieces, it is different from the one presented here: our model uses distributional information between morphs to find out how many morphemes they encode, that is, to actually help the learner in acquiring meaning. For instance, imagine a situation where the learner has detected that a morph$_a$ does not show a systematic syntagmatic relationship with another morph. When the learner assigns a particular morpheme to morph$_a$, he will infer that this is all the semantic content that morph$_a$ encodes, and he will move to another morph.

It would be interesting to investigate how the use of this particular distributional information is managed by the learner when it is combined with the construction of paradigms, since the overall system presented in this chapter would be notably optimised. Let us take for example the scenario that has just been sketched: the learner has detected that a morph$_a$ do not show a systematic syntagmatic relationship with another morph and, once a particular morpheme is assigned to it, for instance nominative, the learner infers that this is all the semantic content that morph$_a$ encodes. Now, let us imagine that the learner also discovers that morph$_a$ is found in the following word-specific paradigm:

Nominative	Xa
Accusative	Xb
Genitive	Xc
Dative	Xd
Ablative	Xe

Table 4.75: The use of the systematic syntagmatic distribution (I)

Once the stem is segmented from affixes, the learner will build the following general paradigm:

Nominative	a
Accusative	b
Genitive	c
Dative	d
Ablative	e

Table 4.76: The use of the systematic syntagmatic distribution (II)

At this point, the learner would not need to find out whether or not another dimension must be included in the paradigm, since he would know that paradigm 4.76 is complete. The reason is that he would extend to all the forms of the paradigm the information regarding the systematic syntagmatic distribution of morph$_a$, namely, the information that points to the non-synthetic nature of morph$_a$. Accordingly, the learner would know by extension that the other forms of the paradigm $\{b, ..., e\}$ are agglutinative and none of the them encodes any other distinction.

Therefore, the line of research presented in this chapter is an interesting step forward in the general enterprise of designing a discovery procedure for language acquisition where abstract linguistic levels are grounded on perceptible levels.

5

The Chunking Procedure

The aim of this chapter, which is based on Fasanella and Fortuny (2012), is to propose a mechanism of morphophonological analysis, the Chunking Procedure, and to relate the acquisition of morphosyntactic properties to the analyses obtained by the Chunking Procedure by means of bootstrapping mechanisms. In proposing a model like the present one, the acquisition of some high-order morphosyntactic patterns is triggered by the learner's analyses of two features relative to morphs: their [±boundedness] and their [±syntheticity]. It will be argued that these features, which must be determined for every morph during language acquisition independently of the postulation of the Chunking Procedure, can be used to trigger the acquisition of certain morphosyntactic properties, those that traditional parameters compress.

Using this methodology we can avoid the problems faced by parametric approaches as well as those problems, noted in chapter 3, related to models that develop algorithms for parameter setting. This proposal allows us to examine patterns of morphosyntactic variation among natural languages in the same terms as those used by the learner when examining his input language. In this sense, this chapter tries to offer another concrete instance of how Greenberg's problem can be derived from Plato's problem.

This chapter is organised in the following fashion: in section 5.1 I will propose the morphophonological mechanism of data analysis that is argued to be used by the learner when acquiring his input language, the Chunking

Procedure. In section 5.2 I will make explicit the bootstrapping mechanisms related to the results obtained by the learner using the Chunking Procedure. In section 5.3 I will develop this argument by illustrating the dynamics of the bootstrapping mechanisms in dealing with the acquisition of three particular sets of related properties: those associated with the so-called Polysynthesis Parameter (5.3.1), the properties associated with pronouns and dropped arguments, as shown by Neeleman and Szendrői (2007) (5.3.2), and the properties associated with languages classified as satellite-framed languages (5.3.3). In section 5.4 I will make some suggestions regarding how linguistic theory could explain that the analyses made by the Chunking Procedure correlate with some linguistic properties, that is, where bootstrapping correlations come from. Section 5.5 is the conclusion of this chapter.

5.1 A mechanism of morphophonological analysis: the Chunking Procedure

In this section a mechanism of morphophonological analysis used by the LAD in order to attain a morphological analysis of the PLD is proposed. As will be argued, the mechanism relates some morphophonological analyses, postulated on independent grounds in language learning, with bootstrapping mechanisms of high-order morphosyntactic properties of the target language, namely, those properties that traditional parameters range over. Therefore, this proposal allows us to base the acquisition of some abstract linguistic phenomena upon the fixation of more perceptible points of variation, with the final aim of developing a grounded theory of language acquisition. By using this methodology, (the resulting) linguistic variation can be examined by the linguist in the very same terms as those arguably used by the learner when analysing the PLD.

Given this objective I will explore the possibility that genuine parameters (points of variation among languages), that is, those satisfying the Atomicity Condition, the Accessibility Condition and the Positive Evidence Condition which we referred to in chapter 3, can be reduced to or expressed in terms

of mechanisms of data analysis independently postulated in acquisition. The mechanism of morphophonological analysis that will be proposed, thus, can be understood as a possible solution to the Locality Problem, the Linking Problem and the Indetermination Problem.

As we argued in chapter 3, in solving these three problems by meeting the three related conditions, we also give a solution to the problems commonly found in learnability models, that is, the Mixed Patterns Problem, the Morphosyntactic Correlations Problem and the Subset/Superset Problem.

I repeat here all these problems and conditions pointed out in chapter 3. Each learnability problem appears before the variation problem related to it, and after them, their respective learnability condition is introduced (see also diagrams in tables 3.1, 3.2 and 3.3 in chapter 3 on the relationship between problems and conditions):

(1) *Locality Problem*
In order to fix the value of a macroparameter the learner should analyse the data he receives in a global and transverse way, since macroparameters are defined on highly general properties that are spread across the target language.

(2) *Mixed Patterns Problem*
Models assuming (macro-)parameters, with associated clustering properties, that apply to the whole language as a single unit or block will not successfully be employed to deal with real natural languages, which very often show mixed patterns.

(3) *Atomicity Condition*
Parameters must be atomic, they cannot be clusters of properties.

(4) *Linking Problem*
Parameters are defined over abstract linguistic entities such as verbs, nouns and pronouns, so the infant still faces the problem of linking these abstract mental representations to actual physical entities in the speech signal (Gervain and Mehler 2010, 194).

(5) *Morphosyntactic Correlations Problem*
Models assuming syntactic parameters codified in UG cannot directly explain why some morphological properties robustly correlate with syntactic properties.

(6) *Accessibility Condition*
Parameters must be set by directly inspecting phonological and morphological properties of utterances.

(7) *Indetermination Problem*
One of the two values of certain parameters cannot be determined on the basis of positive evidence.

(8) *Subset/Superset Problem*
Some learnability models do not take into consideration languages that show subset/superset relations.

(9) *Positive Evidence Condition*
Both values of a parameter must be set on the basis of positive evidence.

First, I will not consider points of variation among languages as options coded in UG that compress multiple and relatively diverse aspects of the morphosyntax of possible natural languages, thereby differing from macroparametric approaches, in order to meet condition (3). Second, the properties to be set by the mechanism of data analysis are not only atomic but can also be set by inspecting phonological and morphological properties of utterances, thereby differing from the higher-order points of variation postulated by microparametric proposals, which are more remote from the morphophonological inspection of the PLD. In following this learnability consideration we follow condition (6). Third, all values to be fixed by the mechanism of data analysis could be set on the basis of positive evidence, meeting condition (9).

A recurrent topic in the acquisitional literature is how infants acquire words in order to continue with the acquisition of more abstract entities

(Gleitman and Fisher 2005). It has been shown that infants are able to detect and extract abstract regularities from the input (Marcus et al. 1999) and it has also been argued that they are capable of using general-purpose statistical methods of analysis (Saffran et al. 1996) as well as phonological cues to segment continuous speech in terms of discrete units (Christophe et al. 1994, Sansavini et al. 1997). However, as objected to by Yang (2004), statistical learning cannot reliably be used to segment words when scaled to a realistic setting. Yang discusses a series of computational models tested on a random sample of child-directed English sentences from CHILDES database. The computational model using only statistical learning methods based on local minima yields poor results, even assuming that the learner has syllabified the input correctly, a non-trivial task. Unsurprisingly, a second model that incorporates a small amount of prior knowledge about phonological structures to statistical learning obtains much better results. The third model considered by Yang, which obtains better results than the previous two models, does not use statistical learning minima whatsoever: it simply stores previously extracted words in the memory to bootstrap new words.

I will, however, leave aside these types of mechanisms based on word acquisition. The reason is that I will not consider the word unit as a primitive of the procedure, given its dubious theoretical and empirical status; thus the present proposal departs on this point from what is standard in the acquisition literature. Actually, most acquisitional studies have been focused on the acquisition of words in English and similar European languages, as Romance or Germanic languages. Nonetheless, the traditional concept of "word" cannot be neatly applied to richly inflected languages, such as Turkish or Georgian. This is the reason why we take the "morph" unit as the primitive of the procedure, whose definition is applicable to all languages.

The minimal morphological category, which can also be called *head*, is understood as proposed in chapter 2:

(10) *Morph*

A linguistic form α, viewed as a string of phonemes, is a morph or

head iff it is meaningful and does not contain any meaningful non-empty proper substring.

The task of segmenting continuous speech into heads or morphs seems to be guided by linguistic knowledge about phonological structures, such as prosodic information, as argued by Kemps et al. (2005). For instance, it has been observed that, in stress-timed languages, a stressed vowel of a noun reduces as a function of the number of unstressed syllables attached to it, as it has been argued for Dutch (Nooteboom 1972), for English (Fowler 1977, Lehiste 1970) and for Swedish (Lindblom and Rapp 1973).[1] This acoustic difference, which is perceived by listeners (Kemps et al. 2005), provides a useful cue to segment complex words into heads.

It seems reasonable to assume that, once a morph has been identified, it is stored in the memory and it is used to bootstrap other morphs, since, as Yang (2004) demonstrates, this mechanism provides the most optimal model to learn non-complex words in a language like English.

We take certain morphophonological properties of heads as a unit of inquiry of the Chunking Procedure. The starting intuition that guides the proposal is that all languages share the same class of grammatical features but differ as to how they realise them morphophonologically. We thus adhere to Cinque's cartographic proposal (Cinque 1999), according to which all languages share the same set (and hierarchy) of grammatical categories, although they may differ precisely as to whether they are realized as free heads, agglutinating suffixes, inflectional suffixes, or non-heads, i.e., adverb phrases.[2]

[1] Kenneth L. Pike proposed in the forties that languages vary in their isochrony, that is, in the rhythmic division of time in equal portions, in three different ways: in stressed-timed languages the duration of stressed syllables is equal, in mora-timed languages the duration of every mora is the same, and in syllable-timed languages every syllable has the same duration. Although this classification is roughly accepted, the empirical validity of these types is still under discussion.

[2] I do not commit myself, though, to the claim that cartographies are primitives of UG, following Fortuny (2008), among other authors.

5.1 A mechanism of morphophonological analysis: the Chunking Procedure

Let us now define the proposed morphophonological mechanism of data analysis:

(11) *Chunking Procedure*

Given a morph *M*, the learner determines:

(a) whether *M* is phonologically dependent of other morphs ([+bound]) or not ([-bound]), and

(b) whether *M* conveys only one morpheme ([-synthetic]) or more than one morpheme ([+synthetic]).

The first property, boundedness, is fixed by the learner by inspecting a string of morphs, and is arguably determined on the basis of phonological cues in the acoustic signal, such as pauses: if a pause can eventually appear before and after a morph, then this morph is unbound; otherwise, the head is bound. Similarly, if a morph appears in isolation it is unbound, and bound otherwise. Language-specific cues may also play a role, such as vowel duration. As Kemps explained, "the observed sensitivity of listeners to these prosodic differences [such as duration and intonation] suggests that these acoustic cues help the perceptual system in determining early in the signal whether an inflected (bisyllabic) or an uninflected (monosyllabic) form is likely to be heard" (Kemps et al. 2005, 46). As noted above, the length of a vowel of a noun in stress-timed languages indicates whether or not the following syllable constitutes a morph bound to this noun. Further investigation is needed, however, in order to determine how the learner accomplishes this mandatory task for the acquisition of morphology.

The second property, syntheticity, is fixed by inspecting how a morph is related to grammatical categories provided by UG, or morphemes. More precisely, the mechanism of data analysis should inspect whether a head conveys a sole morpheme or more. Here not only mechanisms of speech segmentation are involved, but semantic categories provided by UG must also be taken into consideration. In this respect, Pinker (1984)'s model regarding paradigmatic relations, extensively reviewed in the previous chapter, must be

considered. As has been argued, one of the advantages of using paradigms for representing morphological knowledge is that the acquisition of semantic distinctions is simplified: the grammatical information already learned for form of a given paradigm serves as an indexing system capable of attributing this particular information to the novel learned forms that enter into the paradigm. In other words, if the learner has hypothesised that a certain form encodes, for example, number and gender, he will extrapolate this knowledge to all novel forms that will enter into this paradigm, simplifying the whole task of discovering the morphological content of his linguistic input. Apart from that, new information regarding which morphemes are represented by each morph is hypothesised on the basis of phonologically and semantically salient properties of new paradigmatic acquired forms.

Another relevant hint for discovering the syntheticity of morphs is the use of distributional properties in order to find out whether or not a morph shows a systematic syntagmatic relationship with another morph. As argued in chapter 4, this perceptible property correlates with syntheticity: if a morph shows a systematic syntagmatic relationship with another morph it will tend to become fusionalised with it and then it will be [+synthetic], encoding more than one semantic distinction; on the contrary, if a morph does not show a systematic syntagmatic relationship with another morph it will tend to be agglutinative and, thus, [-synthetic].

It seems reasonable to assume that the Chunking Procedure entertains firstly property (a) by considering mechanisms of speech segmentation and secondly property (b) by incorporating semantic considerations as well as distributional information into the morphological analysis, once segmentation has been obtained. Indeed, developmental studies argue for this perspective: children learn to segment morphs, and then use them respecting only one of the several morphematic distinctions they are associated with (say gender), and it is only later that they learn the full set of morphematic distinctions related to these morphs (gender and number, for example), as has been shown by Slobin (1984). These developmenta data suggest, then, that the learner firstly discover the [±boundedness] of morphs, and then their [±syntheticity].

Two clarifications are in order. Firstly, the fact that we take the properties of being [±bound] and [±synthetic] as the basics of the analysis mechanism regarding morphs does not mean that these are the very first operations that the LAD would entertain when trying to learn the target language. Prosodic analyses and the determination of phonological distinctions in the target language, to mention two representative examples, should occur first.

Secondly, the properties of heads of being [±bound] and [±synthetic] must be indispensably fixed by the learner, independently of whether one assumes the mechanism in (11) or not: determining the morphological segmentation and discovering the semantic content of units are mandatory tasks for the learner in order to acquire language, as we already argued in chapter 2 (Clark 2001).

The novelty of this approach, therefore, is to link two necessary morphophonological valuations, independently postulated, and also to relate these valuations with bootstrapping mechanisms of high-order syntactic properties, as will be shown in the next sections.

Let us highlight some general advantages of introducing the Chunking Procedure (11) into the study of language acquisition and language variation. Firstly, the procedure satisfies the above-mentioned intuitive learnability conditions, thereby differing from standard (macro-/micro-) parametric approaches, and also from the formal studies on parameter setting reviewed in chapter 3. The properties of being [±bound] and [±synthetic] satisfy the Atomicity Condition, since they do not range over a variety of morphophonological properties but only over a single morphophonological property: being bound or unbound and conveying a sole morpheme or more, respectively. The properties of being [±bound] and [±synthetic] also satisfy the Accessibility Condition: the learner can detect whether a head is bound or not by inspecting acoustic and phonological properties of utterances, and he can also detect whether a head is synthetic or not by determining, on the basis of a distributional analysis, whether it conveys a sole morpheme or more. This last point is relevant in order to understand what makes the Chunking Procedure different from some microparametric proposals (as discussed in chapter 3): the primitives of the analysis mechanism proposed are detectable

on the basis of PLD, whereas the primitives used in microparametric proposals are too abstract to be detected in the input. Finally the procedure in (11) satisfies the Positive Evidence Condition by offering values that can be fixed through positive evidence in the PLD.

Secondly, within this model, linguistic variation can be attributed to accidental properties of morphs concerning how languages morphophonologically realise and pack features in their linguistic structures. In this sense the proposal we are developing emphasises the minimalist Uniformity Hypothesis, introduced in chapters 2 and 3, and reinforces the understanding of syntax as an invariant component.

Thirdly, patterns of morphosyntactic variation can be derived from the valuation attained by the learner using the Chunking Procedure during the process of language acquisition. As I will illustrate in the next section, the morphophonological valuation attained in this procedure heads towards the acquisition of certain syntactic patterns; thereby it is possible to account for standard cases of linguistic variation as well as for mixed patterns, which cannot be so elegantly analysed in classical parametric terms.

5.2 Bootstrapping mechanisms

The Chunking Procedure is a mechanism that explores the PLD in order to determine how morphs are morphophonologically realised; it is thus part of the morphological analyser of the LAD.

In this section I will show how to use the analysis obtained by the learner with the Chunking Procedure as a trigger for bootstrapping syntactic properties of the target language. I adhere to the following definition, standard in the acquisitional literature, of bootstrapping mechanisms:

(12) Bootstrapping mechanism
Heuristic learning mechanisms that exploit the universal correlations that exist between perceptually available, surface characteristics of a language and its abstract morphosyntactic properties
(Gervain and Mehler 2010, 194).

I will capitalise on the observation that there exist general correlations between abstract syntactic patterns and the analysis obtained by the Chunking Procedure. These correlations reframe observations that have been noted in the literature, as will become clear. I will directly formulate these correlations as bootstrapping mechanisms that link syntactic patterns to the properties set by the Chunking Procedure, before discussing them in turn in the following sections:

(13) *Bootstrapping mechanisms triggered by the Chunking Procedure*

(a) Once the learner has determined that there is a [+bound] morph instantiating a morpheme F, then he can infer that the maximal projection instantiating F in the target language:

- has a free distribution, and
- can be omitted.

(b) Once the learner has determined that there is a [+bound] morph conveying *case* or *number* on pronouns, then he can infer that any argument of the verb can be omitted in the target language.

(c) Once the learner has determined that there is a [-bound] or a [+bound, -synthetic] morph expressing *path*, then he can infer that multiple constructions that are related to the separate lexicalisation of this morpheme are available in the target language.

Someone who is not very familiar with the postulation of bootstrapping mechanisms as heuristic methods to be used by the learner in language acquisition could wonder what differentiates a model like that from the postulation of a set of triggers related to the valuation of *a priori* codified parameters. As Sakas and Fodor (2001) explain, in a process-neutral scenario where triggers or cues are postulated: a) a trigger Tv_i^m is taken to be the specific structural property within sentences that the value v_i^m of a parameter p_i is responsible for licensing; and b) an encounter with Tv_i^m in the target language would constitute reliable evidence for v_i^m in the target grammar.

As the authors explain, in a context of parametric valuation by means of triggers, what remains to be addressed are, at least, the following three questions "whether the learning device can recognise this evidence, whether it adopts v_i^m, and if so by what mechanism" (Sakas and Fodor 2001, 175).

Adopting a model that exploits universal correlations between surface perceptible properties and high-order morphosyntactic properties, instead of triggers for values of parameters, the three open questions noted by Sakas and Fodor (2001) can be addressed: first, the LAD can recognise the evidence because, as is indicated here, it is perceptible; second, once the evidence is recognised, the learner acquires the high-order properties by means of bootstrapping mechanisms, which by definition are automatic or computation-free; third, the mechanism that carries out the process can be argued to be of the sort of the Chunking Procedure.

In the next section it will be shown how certain well-accepted correlations between morphophonological and syntactic properties can be recast in terms of the three heuristic learning mechanisms in (13).

Before concluding this part, note that, on the one hand, (13) (a) is a general heuristic mechanism that may be relative to any grammatical feature: besides our case study of this general mechanism in section 5.3.1, all variation concerning the functional IP-area can be recast in terms of whether a particular morpheme is realised through a maximal projection (an adverb) or a functional head, following Cinque (1999). On the other hand, mechanisms (13) (b) and (13) (c) are relative to particular morphemes: *case* and *number*, and *path*, respectively.

5.3 Deriving clusters of linguistic properties

In this section I focus on how the learner can use the analyses obtained by the Chunking Procedure in order to infer high-order syntactic properties of his target language. Section 5.3.1 is devoted to some of the properties associated with the positive fixation of the Polysynthesis Parameter (Baker 1996), where a brief critical review of that parameter and its formulation is also included. In 5.3.1.1 some Spanish data already presented in chapter 3 are recalled in

order to illustrate how the proposal made in this chapter can successfully deal with mixed patterns. Section 5.3.2 is dedicated to the cross-linguistic correlation between agglutinative morphology on pronouns and radical pro-drop phenomena pointed out by Neeleman and Szendrői (2007) and how it interacts with the bootstrapping mechanism triggered by the analysis of the Chunking Procedure. In section 5.3.3 Talmy's famous classification of languages regarding the expression of the *path* morpheme, the distinction between verb-framed and satellite-framed languages (Talmy 1985, 2000), is examined in relation to the bootstrapping mechanism proposed and the linguistic properties associated to satellite-framed languages.

5.3.1 The Polysynthesis Parameter

A starting general question in Baker (1996)'s work on polysynthesis is the nature and extent of morphosyntactic variation. Two opposite views on the superficial divergences between two apparently very different languages are considered: English and Mohawk. One possibility is that these two languages seem so different from each other as a consequence of a cumulative effect of a plenitude of minor differences. A further possibility is that English and Mohawk differ essentially as to a unique characteristic deeply embedded in their respective grammars that pervades a multitude of linguistic constructions. Baker's work is known to adhere to this second view, that is, to the hypothesis that English and Mohawk look so different because they each have a different "structural genius". This approach leads Baker to the formulation and study of the Polysynthesis Parameter, repeated below:

(14) *The Polysynthesis Parameter*

> Every argument of a head element must be related to a morpheme in the word containing that head.
> Yes: Mohawk, Nahuatl, Mayali, etc.
> No: English, French, Chichewa, etc.

In this formulation a head is understood as an X^0 category in the X-bar format that is associated with an argument structure in the lexicon, and

the morphemes under consideration in the definition of the parameter are agreement morphs.

We must recall that Baker developed the informal formulation of the Polysynthesis Parameter given in (14) into what he considered to be a 'precise principle' concerning θ-role assignment (Baker 1996, 16). This principle was defined as a visibility condition:

(15) *The Morphological Visibility Condition* (MVC)
A phrase X is visible for θ-role assignment from a head Y only if it is coindexed with a morpheme in the word containing Y via:

(a) an agreement relationship with a pronominal affix

(b) a movement relationship (or noun-incorporation).
(Baker 1996, 17)

The two types of coindexing relationships, (a) and (b), were permitted in the standard Principles and Parameters framework wherein Baker's work was developed: (a) and (b) involve, respectively, coindexing between a DP and an agreement category, and coindexing between a moved element and its c-commanded trace in a thematic position, assuming Baker (1988)'s analysis of noun-incorporation in terms of head-movement.

Some languages, like Mohawk, resort to both coindexing relationships, because they display agreement affixes on the verb for several arguments and productive noun-incorporation into the verb; whereas other languages, like Navaho or Warlpiri, show only (a), since they display several pronominal affixes and lack productive noun-incorporation. Other languages do not resort to morphological visibility to make arguments visible for θ-assignment; these languages, for which neither (a) nor (b) holds, are arguably specified with a negative value for the Polysynthesis Parameter.

Although I follow Baker's intuition that so-called non-configurationality and incorporation may be somehow related to the same mechanism,[3] I differ

[3]The label 'non-configurationality' is used as a descriptive tool to refer to the availability of different orders among phrases, but I do not commit myself to a non-configurational analysis of languages displaying these phenomena. See, with regard to this, section 5.4.1.

from his technical implementation for three reasons. One reason is that the MVC does not seem to be a precise principle that appropriately develops the Polysynthesis Parameter. Note that, given the formulation of the MVC, phrases in a non-polysynthetic language like English would be invisible for θ-assignment; accordingly, they would not receive a θ-role and they should be ruled out as impossible languages by the Theta-Criterion:

(16) The Theta Criterion
Each argument bears one and only one θ-role, and each θ-role is assigned to one and only one argument.
(Chomsky 1981, 36)

In other words, the MVC does not distinguish polysynthetic languages from non-polysynthetic languages, but languages for which both (a) and (b) in (15) hold from those from which only (a) holds. Consequently parameter (14) is not properly developed into the principle MVC, but rather the MVC is a subparameter embedded into the Polysynthesis Parameter that sets apart two sets of languages.

Another possible shortcoming is that the Polysynthesis Parameter is related to every θ-role of a language, whereas the MVC is a condition referred to single phrases.

Indeed, if the Polysynthesis Parameter were 'developed' into the MVC, as Baker proposes, then the Polysynthesis Parameter is not a macroparameter of UG, but a byproduct of the above condition for θ-role assignment: the difference between English and Mohawk would derive from a cumulative effect, and not from a principle deeply embedded into each language's grammar, a conclusion that is in contradiction to Baker's position.

Apart from this, the MVC seems unprincipled or hard to motivate on independent grounds. It remains unclear why a DP should need to be coindexed with an agreement affix to receive a θ-role from the verb. Given current standard assumptions on the design of grammar and the locus of variation, it is dubious that morphological considerations must intervene in the licensing of deep syntactic/semantic operations, such as θ-assignment. Instead, I will adhere to the rationale that linguistic variation affects only surface

properties (say, morphophonological properties), and does not involve operations taking place in the Computational and/or in the Conceptual-Intentional components. A more straightforward account can be attained indeed if we think that agreement affixes directly receive a θ-role, an option noted but not followed by Baker (1996) (see section 5.4.1 regarding this option).

Let us proceed to illustrate how the Chunking Procedure may be used to shed light on the problem of how the learner infers syntactic properties of the target language from a morphophonological analysis. Assume that, given an amount of linguistic input, the Chunking Procedure has determined that there is a [+bound] morph M_1 (attached to the verb) that instantiates a particular θ-role θ_1. The learner should be able to determine on independent grounds whether M_1 is an incorporated noun or an affix agreeing with a DP. More precisely, if M_1 can also appear without being incorporated and as a fragment, then it will be a noun, whereas if M_1 is always bound (i.e., it cannot appear freely or as a fragment), then it will be an affix. Consider now the latter situation, in which M_1 is an affix agreeing with a maximal projection. In virtue of the bootstrapping mechanism formulated above in (13) (a), it follows that the maximal projection instantiating the same morpheme$_1$ can be omitted and can have a relatively free distribution.

We can exemplify these dynamics with the case of Basque verbs. In Basque, auxiliary and synthetic verbs agree with the ergative argument, the absolutive argument and the dative argument of the sentence by means of dedicated morphs that always display the same order with respect to the verb. The maximal projections instantiating the ergative case, the absolutive case and the dative case can be omitted when they are topics, and do not have a fixed position in ordinary sentences.[4]

Let us illustrate this pattern with the following Basque examples:

[4]As explained in De Rijk (2007), emphatic topic pronouns and pronouns in focus cannot be omitted, "but keep their position in front of the sentence or verb complex" (De Rijk 2007, 205).

(17) Ni-k aita-ri diru-a eska-tu
I-ERG father-DAT.sg money-ABS.sg ask-prf
d-io-t
Abs.3-(PRES.trans.aux)-DAT.3sg-ERG.1sg
'I have asked father for (some) money.'

(18) Zu-k aita-ri diru-a eska-tu
You-ERG father-DAT.sg money-ABS.sg ask-PERF
d-io-zu
ABS.3-(PRES.trans.aux)-DAT.3sg-ERG.2sg
'You have asked father for (some) money.'

(19) Ni-k zu-ri diru-a eska-tu
I-ERG you-DAT money-ABS.sg ask-PERF
d-izu-t
ABS.3-(PRES.trans.aux)-DAT.2sg-ERG.1sg
'I have asked you for (some) money.'
(Saltarelli et al. 1988, 238)

The Chunking Procedure would analyse the morphs 'd', 't', 'zu', 'io' and 'izu' as [+bound], since it is the case that they always appear in the same order jointly with the verbal auxiliary of non-synthetic verbs. In the examples, above, however, the present tense marker of the auxiliary is deleted due to phonological reasons (Saltarelli et al. 1988, 222). This fact does not affect the proposal, since agreement categories are always bound, independently of whether or not the auxiliary is overtly expressed, and they cannot appear in a [-bound] configuration.

The morph 'd' corresponds to the third person absolutive, morphs 't' and 'zu' are the ergative markers for the first and second person, and 'io' and 'izu' are the dative markers for the third and second person. The learner would infer, following bootstrapping (13) (a), that their associated nominal projections do not have a fixed position in the sentence and can be elided.

From a typological perspective, if the target language displays several [+bound] morphs M_1, ..., M_n, each receiving a particular θ-role θ_1,..., θ_n,

then the language will be described as predominantly polysynthetic. And if each θ-role were assigned to a different [+bound] morph, then the language would be considered purely polysynthetic.

A consequence of the proposed piecemeal procedure is that there is no Polysynthesis Parameter coded in UG, a situation that would also emerge in Baker (1996), as noted, if one relates the informal Polysynthesis Parameter to the precise principle MVC.

5.3.1.1 On mixed patterns

As we mentioned in chapter 3, theories using parameters codified in UG cannot account for mixed patterns. Given such theories it is not expected to find a) languages where some properties of a parametric cluster are found but not others, and neither is it expected to find b) a language where some pieces behave as if the parameter was fixed with a concrete value, but where other equivalent pieces behave in the opposite way, as if the parameter was fixed with the other value. Nonetheless, it is worth keeping in mind that "pure" types of languages are rare, if they exist at all. This issue is developed in Trask (2002)'s review of *The Atoms of Language*, as we already mentioned in chapter 3. In Trask's words, "the existence of (apparently numerous) 'mixed' or 'compound' languages, with complicated combinations of properties that seemingly cannot be neatly fitted into any sets of parameters at all, represents an enormous obstacle for B[aker]'s parametric account" (Trask 2002, 78).

In the argument we are developing, there is no parameter that sets the polysynthetic typological class of a language. That a language is predominantly polysynthetic is a cumulative effect of most of its functional morphs being analysed as [+bound]. This directly predicts the existence of mixed or non-pure languages, i.e., languages that display polysynthetic effects only to some extent. We can recall the Spanish data on clitic doubling presented in chapter 3 regarding mixed patterns. In certain varieties of Spanish, the dative clitic is mandatory, whereas the correferential prepositional phrase can be omitted and can appear in different positions. The dative clitic, *le*, behaves thus like an agreement affix of a polysynthetic language, as Kayne

(2005b), among others, has argued. However, Spanish is not a predominantly polysynthetic language:

(20) Le pedí que viniera
Clitic.DAT.3sg asked that come.SUBJUNC.3sg
'(I) asked him to come.'

(21) Le pedí a Juan que viniera
Clitic.DAT.3sg asked to John that come.SUBJUNC.3sg
'(I) asked him John to come.'

(22) A Juan le pedí que viniera
To John clitic.DAT.3sg asked that come.SUBJUNC.3sg
'To John (I) asked him to come.'

(23) *Pedí a Juan que viniera
Asked to John that come.SUBJUNC.3sg
'(I) asked John to come.'

As is observed in the examples, in these Spanish varieties the dative clitic is mandatory and the prepositional phrase, *a Juan* ('to John'), can be omitted and does not have a fixed position in the sentence.

The global two-valued Polysynthesis Parameter plays no role in the description of this simple pattern. One could be tempted to postulate a minor parameter relative solely to goal arguments, according to which the goal argument can be either realised by a dative affix, as in Spanish, or by a prepositional phrase; similarly, we could postulate a parameter relative to patient arguments, and another relative to locative arguments, and so forth, but this would be no more than an unprincipled way of providing uninteresting descriptions.

The pattern illustrated above can be more neatly explained if we assume a mechanism like the Chunking Procedure which explores the PLD and uses the resulting information to trigger bootstrapping. In this case, the relevant

morph, the dative *le*, is analysed as [+bound] and the availability of the omission of the prepositional phrase and its free distribution follow from the bootstrapping mechanism in (13) (a).

Consequently, it seems that exploring mechanisms of morphophonological analysis in combination of bootstrapping mechanisms may provide a more principled and restrictive way to also account for linguistic variation patterns.

5.3.2 Pronouns and dropped arguments

In this subsection I will be concerned with Neeleman and Szendrői (2007)'s proposal relating the morphology of pronouns with radical pro-drop phenomena, and I will try to derive some of their observations from the analysis obtained by the Chunking Procedure.

Radical pro-drop languages, also known as 'discourse pro-drop languages' (Roberts and Holmberg 2010, 8), allow the omission of nominal arguments as subjects or objects, and also the omission of possessors.

Neeleman and Szendrői's study examines the relation between the morphology of pronouns and the possibility of having or not radical pro-drop; they propose what they call the Radical Pro-Drop Generalisation, which states that radical pro-drop requires agglutinating morphology on pronouns.

The technical machinery they use in order to derive this generalisation is related to some independent assumptions, namely, that null arguments are regular pronouns that cannot be spelled out at PF, that spell-out rules for pronouns target both terminal and non-terminal nodes and, finally, that the Elsewhere Principle regulates competing spell-out rules (see also section 5.4.2 for more details of their proposal).

Let us pay close attention to the predictions made by Neeleman and Szendrői (2007). As the authors claim, the proposed correlation between the form of pronouns and the availability of radical pro-drop makes two predictions. One is the weak prediction, which states that if a language has invariant or fusional morphology on pronouns, then it lacks radical pro-drop phenomena. The other is the strong prediction, according to which if a

language shows radical pro-drop phenomena, then it necessarily has agglutinative morphology for case and/or number on pronouns.[5]

Regarding the weak prediction, Neeleman and Szendrői examine languages with fusional or invariant pronominal morphology, including classical pro-drop languages as well as creole languages. Their sample includes Swedish, Dutch, Afrikaans, Italian, Pashto, Greek, Papiamentu or Tok Pisin, and their prediction is confirmed in all of them: these languages have invariant or fusional morphology on pronouns and, as expected, do not show radical pro-drop.

For the sake of concreteness, let us consider the case of Swedish. Its pronominal paradigm is shown in the following table:

	Nominative	Accusative	Possessive
1 sg	jag	mig	min
2 sg	du	dig	din
3 sg M	han	honom	hans
3 sg F	hon	henne	hennes
1 pl	vi	oss	vår
2 pl	ni	er	er
3 pl	de	dem	deras

Table 5.1: Swedish pronominal paradigm

The paradigm in table 5.1 is completely fusional in the sense that case and number suffixes cannot be identified separately from the stem. This does not mean that some patterns could not be extracted, like the string hVn in third person singular forms, for example. However, as the authors note, this is not enough to establish an agglutinating paradigm.

Their weak prediction is borne out, since Swedish does not allow the omission of subjects, objects and possessors, that is, does not display radical pro-drop phenomena.

[5]As Neeleman and Szendrői (2007) explain, this does not imply that languages with fusional pronouns disallow pro-drop in contexts of rich verbal agreement, just that they systematically disallow radical pro-drop.

Let us make a clarification regarding the Chunking Procedure and the weak prediction by Neeleman and Szendrői (2007). The learner, obviously, needs to discover the full grammatical content of morphs. Accordingly, the LAD, by means of the Chunking Procedure, should be able to determine, given the relevant amount of linguistic data, that pronominal heads like those listed in table 5.1 are synthetic, because they convey not only personal features but also case and number. However, I will not explore any bootstrapping mechanism to implement Neeleman and Szendrői (2007)'s weak prediction. In other words, no bootstrapping mechanism is going to be proposed such that "If the learner analyses a case morph as [+synthetic] with pronouns, he will infer that the target language does not display radical pro-drop phenomena".

The reason is that, provided that a bootstrapping mechanism is understood as the positive implication that could be established between more easily detectable (morphophonological) properties and the presence of high-order (morphosyntactic) properties, I avoid formulating negative implications, according to which the absence of a syntactic pattern is derived from a morphophonological analysis. In our concrete case study, this means that the learner does not need to infer from the synthetic nature of pronouns that rules involving radical pro-drop are unavailable in the target language. In general, the learner does not need to learn what rules are unavailable in the target language, although they may be available in other natural languages, since it is not necessary to postulate heuristic mechanisms to learn absent options. By restricting bootstrapping mechanism to positive implications the acquisition of syntactic properties is simplified.

This last argument can also be paraphrased in terms of why parameters codified in UG are not part of a plausible model to account for language acquisition: the initial state of the Language Faculty does not specify the set of syntactic properties of all possible languages, or, in other words, a particular I-language is not attained by selecting which of these properties are present in the target language.

With respect to the strong prediction, according to which all radical pro-drop languages have agglutinative morphology on pronouns, Neeleman and

Szendrői consider a great number of typologically unrelated languages as Korean, Burnese, Assamese, Hindi/Urdu, Chinese, Kokota, Cheke Holo and Turkish. Leaving aside the complexities in the analyses of the pronominal paradigms of these languages, the strong prediction is also confirmed with respect to the sample, although Finnish seems to be a counterexample (it is agglutinative for case but only allows subject drop in the first and the second person).

We take as an illustration the case of Korean. In this language, pronouns carry the same case particles that nouns do as suffixes. We reproduce below (some of) the pronominal stems in table 5.2 and the case endings they could be related to in (24):

	1 p	2 p	3 p
Sg	/na/	/ne/	/ku/
Pl	/wuli/	/ne-huy/	/ku tul/

Table 5.2: Korean pronominal stems

(24) Korean case endings

(a) Nom: /ka/

(b) Acc: /(l)ul/

(c) Gen: /uy/

(d) Dat: /ey/; /eykey/; among others.

Combining stems with case endings results in inflected pronouns, like *ku-tul-ka* '3p-pl-Nom'. As expected, Korean allows omission of subjects, objects and possessors, that is, it shows radical pro-drop.

Neeleman and Szendrői (2007)'s observations regarding the strong prediction seem a perfect candidate for postulating a boostrapping mechanism that can be linked to the analysis obtained by the learner using the Chunking Procedure. Assume the learner has detected in the linguistic input that

there is a morph analysed as a [+bound] head instantiating the category of *case* or *number*. Crucially, the relevant morph must be analysed as [+bound] precisely to pronouns to trigger the bootstrapping postulated in (13) (b). This means that LAD should keep track of the grammatical category of lexical units, such as pronouns, a task that is independently motivated for language acquisition and that arguably cooperates with the Chunking Procedure. This may differ from other cases where the analysis obtained by the Chunking Procedure is enough to trigger bootstrapping. To make this point clear, consider for example the case of pronominal clitics. If (pronominal) clitics instantiating a particular θ-role can only be bound to verbs, then analysing the relevant head as [+bound] would be enough to trigger the bootstrapping (13) (a). However, note that *case* or *number* morphs in languages like Korean could appear bound both to nouns and to pronouns, being only in the later case relevant for Neeleman and Szendrői's generalization and for the bootstrapping we are exploring here (for a different perspective on this point that would eliminate the necessity of keeping track of the grammatical category of these pieces in order to trigger the boostrapping in (13) (b), see the proposal by Rubio and Fasanella (2014) sketched in section 5.4.)

Once the relevant morph instantiating *case* or *number* is recognised in the linguistic input and is analysed as [+bound] to pronouns, the LAD follows the bootstrapping mechanism formulated in (13) (b) and infers that the target language allows radical pro-drop, in which case verbal arguments and possessors can be phonologically omitted.

5.3.3 The satellite-framed/verb-framed distinction

In this section I will consider Talmy (1985, 2000)'s famous division between satellite-framed languages and verb-framed languages. In his influential work, Talmy defines a typological distinction between languages depending on how they express the morpheme *path* in events of motion:

> Path appears in the verb root in 'verb-framed' languages such as Spanish, and it appears in the satellite in 'satellite-framed' languages such as English (Talmy 2000, 117-8).

The postulated distinction can be observed in the following examples in English and Spanish:

(25) The rock rolled down the hill
(Talmy 2000, 29)

(26) La botella salió de la cueva (flotando)
the bottle moved.out from the cave (floating)
'The bottle floated out of the cave.'
(Talmy 2000, 49)

In example (25) the verb 'roll' expresses motion and manner of motion and it is the preposition 'down' what encodes the morpheme *path*. However in example (26) the verb *salir* ('to go out') expresses both motion and *path*, whereas manner must be lexicalised in a different word, in this case in the gerund *flotando* ('floating'). Following this criterion English is classified as a satellite-framed language and Spanish, as a verb-framed language.

Acedo-Matellán (2010) argues that satellite-framed languages must be subdivided into two types according to how they morphophonologically realise *path* in the satellite of the verb. Following his terminology, in strong satellite-framed languages *path* is not affixed to the verb, but is expressed in another word, like in English, Dutch, Hungarian or Finnish. Weak satellite-framed languages show *path* affixed to the verb, like in Latin, Slavic or Ancient Greek. Therefore, according to Acedo-Matellán (2010)'s proposal, three linguistic types arise from the morphophonological realisation of *path* in natural languages, namely, verb-framed languages on one side and strong satellite-framed languages and weak satellite-framed languages on the other:[6]

[6]As noted by Real-Puigdollers (2013), these three linguistic types are parallel to the unergative verb types defined by Hale and Keyser (1998):

(a) Catalan: *parlar* ('speak'), English: 'to speak'.

(b) Basque: *lo egin* ('sleep do'), English: 'to sleep'.

(c) Jemez: *sae'a* ('work-do'), English: 'to work'.

(27) En Joan sortí
the Joan go.out.PAST.3sg
'Joan went out.'

(28) John danced into the room

(29) Flatus arbusta e-volvens
gust.NOM.masc shrub.ACC.pl out-roll.PTCP.PRS.NOM.sg
'A gust of wind rolling shrubs out.'
Lucr. 6, 141 (Acedo-Matellán 2010, 97)

Sentence (27) is an example of a classical verb-framed language, Catalan, where *path* is expressed within the verb as usual. In examples (28) and (29) we could appreciate the difference between a strong satellite-framed language, with *path* expressed by a phonological word different from the verb by means of the strong preposition 'into'; and a weak satellite-framed language as Latin, where *path* is expressed using a prefix 'e' attached to the verb.[7]

What is particularly interesting about Talmy's typological distinction is that it has been proposed that satellite-framed languages display some constructions that are banned in verb-framed languages. Acedo-Matellán (2010) demonstrates that Latin is correctly classified as a (weak) satellite-framed language insofar as it shows these constructions. We illustrate this phenomenon by considering Complex Directed Motion constructions, Unselected Object constructions and Complex Effected Object constructions, borrowing Acedo-Matellán (2010)'s terminology, in English, Latin and Catalan.

In Complex Directed Motion constructions a directed motion event is expressed by some element in the structure, whereas the verb expresses a manner co-event:

(30) They danced out of the room
(Zubizarreta and Oh 2007, 128)

[7]In Latin, as Acedo-Matellán (2010) argues and exemplifies, *path* is very frequently expressed by means of a verbal prefix, but it could also be expressed through a prepositional phrase, a combination of a prefix and a prepositional phrase, a combination of a prefix and a determiner phrase and, marginally, a case-marked determiner phrase.

(31) Simulatque e navi e-gressus est
as-soon-as out ship.ABL out-walk.PTCP.PRF AUX.3sg
dedit
give.PAST.3sg
'As soon as he walked out of the ship, he handed it over.'
Cic. Verr. 2, 2, 10 (Acedo-Matellán 2010, 111)

(32) *Els nois ballen fora de la cuina
the boys dance.PRES.3pl out of the kitchen
'Boys dance out of the kitchen.'

Whereas in English (30) and Latin (31) Complex Directed Motion constructions are grammatical, in Catalan they are ruled out. The sentence in (32) is only possible in Catalan with a different meaning, namely, 'Boys are dancing outside the kitchen'.

In Unselected Object constructions an internal argument not semantically selected by the verb and not omissible is present, such as 'John' in (33), as well as some predicative element which is the licenser of the unselected argument, as 'deaf' in (33):

(33) Sue shouted John deaf
(McIntyre 2004, 525)

(34) [Serpentes] [ova] solida hauriunt,[...] atque
snake.NOM.masc.pl egg.ACC.pl whole.ACC.pl swallow.3pl and
putamina ex-tussiunt
shell.ACC.pl out-cough.3pl
'Snakes swallow the eggs whole and expel the shells through cough.'
Plin. Nat. 10, 197 (Acedo-Matellán 2010, 122)

(35) *La Sue cridà en John sord
the Sue shout.PAST.3sg to+the John deaf
'Sue shouted John deaf.'
(Acedo-Matellán 2010, 120)

As can be observed, the Catalan counterpart of the Unselected Object construction in English is impossible, whereas Latin does show these type of sentences.

Complex Effected Object constructions are predicates that involve an object interpreted as a created object and a verb that expresses how the event is carried out:

(36) John baked the cake

(37) Qui alteri misceat mulsum
who.NOM another.DAT mix.SUBJUNC.3sg (honeyed)wine.ACC
'He who makes honeyed wine for someone else.'
Cic. Fin. 2, 5, 17 (Acedo-Matellán 2010, 149)

(38) *En Joan fornejà el pastís
the Joan bake.PAST.3sg the cake
'Joan baked the cake.'

Sentences in (36) and (37) express predicates of object creation, whereas the sentence in Catalan (38) is only possible with a change-of-state interpretation, not with a creation reading.

According to these data, it seems that Talmy's initial intuition is on the right track: when the morpheme *path* is realised not synthetically with the verbal stem (as in satellite-framed languages), then all these types of complex constructions are available. In other words, it seems that the lexicalisation of the morpheme *path* separately from the verbal stem is what enables the presence of the relevant constructions.

The analysis obtained by the Chunking Procedure could be used to bootstrap precisely this availability in the target language. Let us consider the three relevant possibilities. Assume that the Chunking Procedure has detected a morph M_1 expressing solely *path*; then there are two subcases: M_1 has been analysed as [-bound] if the target language is a strong satellite-framed language, like English, or M_1 has been analysed as [+bound, -synthetic] if the target language is a weak satellite-framed language, like Latin. In both cases, given the bootstrapping mechanism defined in (13) (c), the LAD infers the availability of the relevant set of constructions.

Consider now the third possibility, where M_1 is a head synthetically expressing *path* and further morphemes, such as motion; in this case, the target

language would be a verb-framed language, like Catalan. The LAD must conclude that the head 'sort' (the verbal stem of *sortir*, 'to go out') is [+synthetic] insofar as it expresses more than one morpheme, but we do not derive bootstrapping mechanisms from this analysis for the same reason discussed in the previous section concerning negative implications: only the presence (and not the absence) of syntactic properties are derived by the proposed bootstrapping mechanisms. This means that, if the target language is a verb-framed language, the learner does not infer that the target language does not have the set of constructions under discussion, but simply it does not develop the possibility of generating them, whence their ungrammaticality.

5.4 Towards theorical explanations of bootstrapping correlations

In this section I will sketch some theoretical proposals available in the literature that can directly or indirectly account for why it is the case that the three proposed bootstrapping mechanisms in (13) correlate with the morphophonological analyses obtained by the Chunking Procedure.

It is worth noting that this section is addressed to the question of how linguistic theories account for specific phenomena, namely, the cross-linguistic correlation between the morphological and the syntactic properties considered by the bootstrapping mechanisms in (13). In the last sections, we have proposed that there exists some universals correlations that can be exploited by the learner between accessible properties and less perceptible properties in natural languages, and that these correlations define patterns of variation; we have made explicit these perceptual features as well as the related bootstrapping mechanisms. Accordingly, the proposal I have presented in this chapter is not contingent on the particular technical explanations that aim to explain why these correlations hold given a concrete theoretical framework.

In what follows, I will briefly outline three explanations, some of them more commonly accepted than others, for each of the bootstrapping mechanisms proposed in (13).

5.4.1 Polysynthetic behaviour

Let us briefly summarise a proposal in the literature, namely, Jelinek (1984), in order to exemplify a theoretical explanation capable of deriving why the bootstrapping mechanism in (13) (a), repeated below, holds.

(39) Once the learner has determined that there is a [+bound] morph instantiating a morpheme *F*, then he can infer that the maximal projection instantiating *F* in the target language:

- has a free distribution, and

- can be omitted.

As Baker (1996) notes, the observation that in many languages a phrase does not have a fixed position in the sentence and can be omitted when there is a mandatory affix bound to the verb instantiating the same semantic content has been observed since, at least, von Humboldt and his analysis of Aztec in the first half of the nineteenth century, as argued by Foley (1991).

Since then, this phenomenon has been explained appealing to the intuitive idea, commonly found in Amerindian linguistics, that in these constructions inflectional affixes count as pronouns in that they provide the true subject and object of the verb. As a consequence of that, full phrases encoding the same semantic information become optional and can be omitted since they "have the status of some kind of adjunct or modifier" (Baker 1996, 11).

This intuitive idea have been formalised in different frameworks and in different periods. Within Generative Grammar, Jelinek (1984) offers an explanation of this correlation within the Principles and Parameters framework. Jelinek (1984) argues that bound grammatical morphemes in polysynthetic structures, such as subject markers and objects markers, are not just grammatical clitic markers, but constituents that contain fully referential pronouns "that serve as verbal arguments and are case-marked, thereby marking grammatical relations" (Jelinek 1984, 76). In her model, bound morphs receive case from the verb and occupy the structural positions dedicated to true arguments, as the subject and the object positions. Actually, Jelinek (1984)

argues that Warlpiri, a language spoken in the Northern area of Australia, has to be considered a configurational language insofar as affixes in Warlpiri do have a fixed position in the verb complex (and also with respect to the noun). This account is known in the literature as the pronominal argument approach (Legate 2002).

Jelinek (1984)'s model can be extended and her insights applied to other languages showing polysynthetic patterns.

Therefore, according to the pronominal argument approach, Warlpiri would only be non-configurational with respect to full phrases, which can appear in more than one position and are optional, since they behave like some kind of adjunt. Actually, Legate (2002) has offered another account for Warlpiri phenomena based on the cumulative effect of microparameters that, though differing from Jelinek (1984)'s proposal, also considers Warlpiri a configurational language.

These type of proposals provide a possible theoretical path to link morphosytactic properties to the morphophonological analysis of morphs. Jelinek (1984)'s proposal implements a technical explanation of why it is the case that the bootstrapping mechanism in (39) holds, which is used by the learner as a heuristic mechanism: when he encounters a morph which is analysed as [+bound], he can infer that the full phrase that implements the same semantic content can be omitted and does not have a fixed position.

5.4.2 Agglutinative pronouns and radical pro-drop

I will review two proposals in the literature, Neeleman and Szendrői (2007) and Rubio and Fasanella (2014), in order to show how the bootstrapping mechanism in (13) (b), repeated in (40), can be derived:

(40) Once the learner has determined that there is a [+bound] head conveying *case* or *number* on pronouns, then he can infer that any argument of the verb can be omitted in the target language.

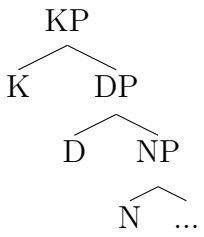
As has been already noticed, Neeleman and Szendrői (2007) appeal to some independent assumptions in order to explain why radical pro-drop phenomenon

cross-linguistically correlates with agglutinative morphology of pronouns. Let us briefly summarise their proposal. Neeleman and Szendrői (2007) argue that pro-drop occurs when regular pronouns in syntax fail to be realised at the PF interface, giving as a result null arguments. Following this logic, the authors propose that radical pro-drop languages would have the following spell-out rule:[8]

(41) $[_{KP}$ +p(ronominal), -a(naphoric)] $\Leftrightarrow \oslash$

The authors assume that the extended nominal projection consists of an NP, dominated by a DP, dominated by a KP:

(42)



The Elsewhere Principle (Kiparsky 1973), assumed by Neeleman and Szendrői (2007), favours the spell-out of a category C over the spell-out of categories contained in C and, also, it gives preference to the phonological realisation of a category C that spells out more features of C over a phonological realisation that spells out less features of C. In a language with fusional morphology on pronous, like Dutch, where pronouns are fusional with respect to case, the existence of a more specific spell-out rule like the one in (43) will block the rule in (41):

(43) $[_{KP}$ +p(ronominal), -a(naphoric), 3, sg, M, ACC] $\Leftrightarrow$ /hem/

Thus in languages whose pronominal paradigm consists of spell-out rules for KP (that is, in languages with fusional pronouns for case and/or number),

[8]Regarding the following notation, the authors explain: "We use the features [+p(ronominal) -a(naphoric)] to indicate that K[ase]P is a pronoun. We are not committed to these particular features. What is important for our account is that pronouns can be distinguished from other nominal categories, such as R-expressions and anaphors" (Neeleman and Szendrői 2007, 682 fn. 7).

a general pro-drop rule, as the one in (41), would not have any effect due to the Elsewhere Principle; actually "its application would be systematically suppressed by the more specific spell-out rules introduced by overt pronouns" (Neeleman and Szendrői 2007, 687).

However, in languages, like Japanese, where case is agglutinative, the setup for overt pronouns and case endings would be the following in (44):

(44) Overt pronouns in Japanese

(a) /watasi/ ⇔ [$_{NP}$ +p(ronominal), -a(naphoric), 1, sg]

(b) /anata/ ⇔ [$_{NP}$ +p(ronominal), -a(naphoric), 2, sg]

(c) /kare/ ⇔ [$_{NP}$ +p(ronominal), -a(naphoric), 3, sg, M]

(d) /kanozyo/ ⇔ [$_{NP}$ +p(ronominal), -a(naphoric), 3, sg, F]

(e) /tati/; /ra/ ⇔ [pl]

(45) Case endings in Japanese

(a) /ga/ ⇔ [$_{K}$ Nom]

(b) /o/ ⇔ [$_{K}$ Acc]

(c) /ni/ ⇔ [$_{K}$ Dat]

(d) /no/ ⇔ [$_{K}$ Gen]

In order to have a pronoun in Japanase, then, one has to combine a [$_{NP}$ +p(ronominal), -a(naphoric), P, N, (G)] piece with the corresponding case ending morph. For instance, /karega/ would be the pronoun for the third person singular masculine in nominative case.

Accordingly, the general zero spell-out rule in (41) is not in an elsewhere relation to any of the rules in (44). As Neeleman and Szendrői (2007) argue, on the one hand, the description in (41) is more specific than those in (44) insofar as the former spells out a larger chunk of structure: the rule in (41)

spells out the full KP, whereas the rules in (44) for pronouns spell out the layer of the NP.

On the other hand, the structural descriptions in (44) contain features, concretely person and number, that the zero spell-out rule in (41) is insensitive to, which makes rules in (44) more specific. It is because of this reason that the domain of application of the zero spell-out rule in (41) does not properly include those of overt pronouns; and the other way around, none of the domains of application of the rules in (44) include that of the zero spell-out rule. Consequently, there do not exist blocking effects between (41) and (44) and the drop of arguments is possible for all pronominal arguments.

As these authors explain, the logic of their proposal can be extrapolated to other agglutinating morphemes realising categories lower than K in the extended nominal projection, as for instance *number*: if a language encodes plurality on pronouns using a separate morph, the pronominal stem cannot correspond to the full KP and, then, their spell out rules will not cause blocking effects with the zero spell-out rule in (41). As a consequence, radical pro-drop will be possible.

Neeleman and Szendrői (2007)'s theory would technically explain, thus, why the syntactic phenomenon of radical pro-drop correlates with a more perceptible property: the agglutinative morphology of pronouns.

Nonetheless, in Rubio and Fasanella (2014) another proposal is put forward regarding why languages allowing radical pro-drop phenomena have agglutinative morphology on pronouns and also typically lack verbal agreement. If this proposal is on the right track, there can be a bootstrapping mechanism that not only correlates the property of agglutination of pronouns with the morphosyntactic property of radical pro-drop, but also with the morphosyntactic property of lack of verbal agreement. Then, the bootstrapping mechanism in (40) could be extended in order to include the property of lacking verbal agreement as well. It is because of this potential improvement of the bootstrapping model presented in this chapter that we are going to briefly summarise the basics of this proposal in what follows.

As has been already pointed out, the main concern of Rubio and Fasanella (2014)'s study is that radical pro drop phenomena has been correlated with

two apparently independent facts: the agglutinative morphology of pronouns, as shown by Neeleman and Szendrői (2007), and the lack of verbal agreement. This last property has been noted by many authors and has been approached from very different perspectives, as for instance in the proposals by Huang (1984), Speas (1994), Saito (2007) or Roberts and Holmberg (2010). Both the agglutinative nature of pronouns and the lack of verbal agreement in radical pro-drop languages has been observed in cross-linguistically unrelated languages like Chinese, Japanese, Malayalam, Turkish, Korean and Hindi/Urdu.

Rubio and Fasanella (2014)'s study tries to offer a unified account of these three phenomena arguing that they all can be derived from a single property of those languages, namely, that they lack true referential pronouns, where referential pronouns include personal pronouns, argumental pronouns and possessors; clitics, on the contrary, are considered to be instances of verbal agreement, following Roberts (2010). Therefore, on the basis of Longobardi (1994, 2013) and Neeleman and Szendrői (2007), Rubio and Fasanella (2014) argue that radical pro-drop languages make use of NPs because they lack referential pronouns.[9]

Assuming that radical pro-drop languages use lexical items that behave like NPS from a purely syntactic point of view, although in some contexts they show a deictic meaning, allows an elegant account of why radical pro-drop is found along with lack of verbal agreement and agglutinative morphology.

First, the lack of verbal agreement in radical pro-drop languages can be explained insofar as NPs can only agree in the third person, leaving aside, of course, cases of differential agreement where the verb agrees in third person although a second person pronoun is present, as in 'you linguists are very happy'. Therefore, if all referential pronouns in the relevant languages are actually NPs, then it is expected that the only kind of person agreement they can enter into is third person agreement. Given that having just one type of

[9]Indeed this is somehow related to the observation that omitted arguments in Japanese, a radical pro-drop language, sometimes behave like elided NPs, rather than like pronouns (Hoji 1998).

verbal agreement is logically the same as not having agreement at all, since any change in verbal arguments is not going to trigger any change in verbal morphology, the lack of verbal agreement in radical pro-drop languages is intuitively derived.

Second, if we take these so-called pronouns in radical pro-drop languages to be NPs, their agglutinative nature can also be derived, on the basis that it is cross-linguistically observed that nominal stems can be more easily segmented from their affixal morphology than fusional referential pronouns, like the ones in Romance languages, for example. Crucially, it is possible to apply the model about the acquisition of morphological paradigms and the Minimise Paradigm Constraint (MPC) presented in chapter 4 in order to derive the agglutinative morphology of so-called pronouns in radical pro-drop languages, since it is cross-linguistically observed that nouns do not show systematic syntagmatic relationships with affixes.

Let us take the case of Japanese. In Japanese, alleged pronouns are not in a systematic syntagmatic relationship with any suffix, as can be observed in the following sentences. In example (46), the "pronominal" stem for the first person singular, *watashi*, and the dative marker, *ni*, appear adjacently, but in (47) another morphological marker, the one for plural *tachi*, appear in between the two. Also, in example (48), an atributive sentence, the "pronoun" appears alone, without any suffix after it:

(46) Ano hito wa watashi ni tegami o okutta
That person Topic I DAT letter ACC send.PAST
'That person sent me a letter.'

(47) Ano hito wa watashitachi ni tegami o okkutta
That person Topic I.pl DAT letter ACC send.past
'That person sent us a letter.'

(48) Otto wa kare desu
My.husband Topic he to.be
'My husband is he.'

It is clear, then, that so-called pronouns in Japanese do not show a systematic syntagmatic relationship with possible affixes.

In fact, "pronominal" stems in Japanese show the same distribution with respect to affixal markers than prototypical nouns; or the other way around, nouns show the same affixal specifications as pronouns in the examples above. This distribution can be considered another argument of why so-called pronouns and nouns in radical pro-drop languages would belong to the same grammatical category. In the following sentences, the prototypical nouns *hito* and *shachou*, 'person' and 'chairperson', are followed by the dative case marker (49), the noun *hito* is followed by the plural marker *tachi* before the case marker (50), and the noun *shachou* appears without any morphological specification (51):

(49) Shachou wa ano hito ni tegami o okutta
Chairperson Topic that person DAT letter ACC send.PAST
'The president of the company sent a letter to that person.'

(50) Sachou wa ano hito tachi ni tegami o okutta
Chairperson Topic that person pl DAT letter ACC send.PAST
'The president of the company sent a letter to those people.'

(51) Otto wa shachou desu
My.husband Topic chairperson to.be
'My husband is the president of the company.'

Therefore, Rubio and Fasanella (2014) can account for the agglutinative nature of "pronominal" affixes in radical pro-drop languages appealing to their non-systematic syntagmatic distribution.

This distribution between "pronominal" stems and affixes in Japanese is very different from the picture shown by true referential pronouns, as those in Romance languages, for instance. In those languages pronouns are fusional with their case and number endings in the sense that pronominal pieces are not separable and, hence, no material can appear in between them. In Catalan, for example, although it is possible to find some regular patterns in the pronominal personal paradigm, as the morph *-s* for plural forms, nothing can appear in between the pronouns and this plural morph, they form an indivisible unit:

1 sg	jo
2 sg	tu
3 sg M	ell
3 sg F	ella
1 pl	nosaltres
2 pl	vosaltres
3 pl M	ells
3 pl F	elles

Table 5.3: Catalan pronominal personal paradigm

Actually, following the logic of Rubio and Fasanella (2014)'s proposal, the radical pro-drop phenomenon itself can be understood as a result of lacking true referential pronouns to fulfill argumental roles. When an argument in a radical pro-drop language is omitted, it can simply be dropped, since these languages do not have the possibility of resorting to pronouns or verbal agreement/clitics.

If the relevant radical pro-drop languages do not have referential pronouns but NPs, then it is expected that these so-called pronouns will admit the kind of modification regular NPs admit. Indeed, this is exactly what happens in Japanese and Malayalam, two of the radical pro-drop languages mainly examined by Rubio and Fasanella (2014). In the examples below, it can be observed that Japanese and Malayalam supposed pronouns can ordinarily receive nominal complementation, unlike what happens with Spanish true referential pronouns. It is important to note that definite articles in the Spanish examples do not influence the grammaticality of the sentences, which are ungrammatical anyway without them.

The following examples illustrate that adjectival modification is possible in Japanese and in Malayalam, but impossible in Spanish:

(52) *El cansado él llegó a la estación
'Tired he arrived at the station.'

Spanish

(53) Utsukushii kanojo wa eki ni tsuita
Beautiful she Topic station to arrived
'Beatiful she arrived at the station.'

Japanese

(54) Avan ksheenichu ethi
He tired arrived
'Tired he arrived'.

Malayalam

Japanese and Malayalam "pronouns" can be modified by a specificative relative clause, contrary to what happens in Spanish:

(55) *El él que vino a la fiesta se casó el mes pasado
'The he who came to the party got married last month.'

Spanish

(56) Paati ni kita kare wa sengetsu kekkonshita
Party to came he Topic last month got married
'He who came to the party got married last month.'

Japanese

(57) Aa partykk vannirunna avan, puthiya car vaangi
That party had come he, new car bought
'He who came to the party bought a new car.'

Malayalam

Note that some non-radical pro-drop languages allow structures of the kind "He who...", "Lui quello...". It seems that these are not counterexamples to this proposal. In the first place, it is relevant that these structures in English or Italian are much better with third-person pronouns, unlike Japanese or Malayalam, which productively admit such modification for all persons. In the second place, sentences with these structures in non-radical pro-drop languages are instances of explicative relative clauses, they cannot be specificatives, which points to their non-referentiality.

In Japanese and Malayalam, modification by a prepositional phrase is also possible, whereas in Spanish this complementation with pronouns is completely ill-formed:

(58) *Los ellos de Canarias
'The they from the Canaries.'

Spanish

(59) Nihon no kare
Japan from he
'He from Japan.'

Japanese

(60) Englandil ninnanullavar
England from they
'They from England.'

Malayalam

Finally, so-called pronouns in Japanese and Malayalam can co-appear with a demonstrative, unlike what happens in Spanish:

(61) *Esa ella
That she

Spanish

(62) Sono kanojo
That she

Japanese

(63) Aa aval
That she

Malayalam

All these processes of modification are productive in Japanese and Malayalam with all alleged pronouns, independently of their person, number, case and other specifications.

Therefore, given Rubio and Fasanella (2014), not only would radical pro-drop correlate with the agglutinative nature of case and/or number specifications on so-called pronouns, but also with the lack of verbal agreement. If this line of research turns out to be correct, the bootstrapping mechanism in (40) could be extended in order to capture this generalisation as well, which would become: once the learner has determined that there is a [+bound] morph conveying *case* or *number*, then he can infer that a) the target languages does not have verbal agreemnet and b) that any argument of the verb can be omitted.[10]

Another potential improvement of Rubio and Fasanella (2014)'s proposal with respect to the bootstrapping mechanism proposed to derive radical pro-drop phenomena is that the Chunking Procedure would not need to be combined with a device for recognising the grammatical categories of linguistic units: as long as the learner encountered a [+bound] *case* morph, he would bootstrap the possibility of omitting subjects, objects and possessors, as well as the lack of verbal morphology (contrary to the picture sketched in section 5.3.2, where the bootstrapping mechanism needs to be combined with a mechanism that recognises that the *case/number* morph is [+bound] specifically to pronouns).

5.4.3 The lexicalisation of path and related constructions

In this part I will adhere to Acedo-Matellán (2010)'s explanation of why verb-framed languages do not display the set of resultative constructions associated with satellite-framed languages, that is, to derive the bootstrapping mechanism in (13) (c), repeated here:

[10]For simplicity, we have not considered how Rubio and Fasanella (2014)can deal with apparently mixed cases: languages such as Latin, for instance, display agglutinative case morphology in NPs, but shows verbal agreement. As the authors explain, these cases can be handled introducing some slight modifications in the bootstrapping mechanism triggered by the analysis obtained by the learner using the Chunking Procedure.

(64) Once the learner has determined that there is a [-bound] or a [+bound, -synthetic] head expressing *path*, then he can infer that multiple constructions that are related to the separate lexicalisation of this head are available in the target language.

The crucial issue in Acedo-Matellán (2010)'s account is related to the morphophonological expression of the *path* morpheme, which needs to be independently lexicalised with respect to the verb in order to enable the appearance of resultatives constructions in examples (30) through (38), in the same line of Talmy's first intuitions.

In Acedo-Matellán (2010)'s study it is argued that cross-linguistic differences among natural languages, as the distinction between languages that lexicalise the *path* morpheme in a satellite or within the verb, derive from purely morphophonological differences in the operations triggered in the PF branch by language-specific morphophonological features of functional items; or, in other words, from language specific packaging mechanisms of morphs and morphemes. Given this view, completely compatible with the ideas put forward in this thesis, the author develops a model which acounts for the main differences between verb-framed and satellite-framed languages.

Assuming the cartography for PF suggested by Embick and Noyer (1999, 2001), Acedo-Matellán (2010) proposes to analyse the head *path* in the following fashion given the three attested cross-linguistic patterns:

(a) In a verb-framed language, like Catalan, the verb head v is specified to be obligatorily fused with *path* into a single head before Spell-Out to the PF branch by means of a fusion operation. Since Fusion applies on sister heads, that is, on heads which form a complex head, a Lowering operation is proposed to bring the v head down to *path* before Fusion occurs. These operations are defined as follows:

(65) *Lowering* takes a head and lowers it to the head of its complement creating a complex head out of two heads.

$[_{XP}\ \mathrm{X}^0 ... [_{YP} \mathrm{Y}^0 ...]] \rightarrow [_{XP} ... [_{YP} ... [_{Y0} \mathrm{Y}^0 + \mathrm{X}^0] ...]]$

(66) *Fusion* takes two single sisters and creates a novel single head out of them. The resulting head retains the features of the fused heads.

$[_X$ X+Y$] \rightarrow$ Z$_{X+Y}$

(b) In a weak satellite-framed language, like Latin, *path* is realised independently from the verb but in the same word due to a Lowering operation. As a result of that, the *path* head gets affixed onto the v head. The verb can anyway enter into an adjunction relation with another root interpreted as a *manner* co-event.

(c) In strong satellite-framed language, like English, the *path* head is expressed in another word, and the functional v can enter into an adjunction relation with a root interpreted as a *manner* co-event.

Assuming this model, in verb-framed languages it is not possible to express a *manner* co-event with the verbal head because v and *path* had already formed a fusioned node, and then the separate root with the manner specification cannot adjoin to v. It is for this reason that in sentences like (26), repeated in (67), the manner co-event is lexicalised in a different word, *flotando* ('floating'):

(67) La botella salió de la cueva (flotando)
the bottle moved.out from the cave (floating)
'The bottle floated out of the cave.'
(Talmy 2000, 49)

In satellite-framed languages, on the contrary, the verbal head can adjoin to the manner co-event since the head *path* is not fused with v, creating structures such the one in (25), repeated in (68), where the v head is adjoined with the manner, giving as a result the form *rolled*:

(68) The rock rolled down the hill
(Talmy 2000, 29)

What is particularly interesting for our purposes is that the Lowering and subsequent Fusion operations that are proposed to happen in verb-framed languages between the *v* head and the *path* head can also account for the lack of resultative constructions in verb-framed languages for the same reason adduced before, as Acedo-Matellán (2010) explains:

> Typical s[atellite]-framed constructions are analysed as involving the adjunction of a root to v, being interpreted as a manner co-event. In v[erb]-framed languages, this adjunction structure is not compatible with the Fusion operation obligatory for v and Path. In fact, Fusion operates only on simple sisters heads, so it cannot apply on a complex head which already includes a complex head" (Acedo-Matellán 2010, 93).

The bootstrapping mechanism proposed in (64) can now be technically accounted for: when the morph encoding *path* is not fused with other morphs, concretely with the *v* root, the set of constructions that depend on it to be possible are available in the language: the *v* root can enter into an adjunction relation with another verbal root. However, when the morpheme *path* cannot be isolated because it is expressed with the verbal morph in a single piece, the resultative constructions cannot be formed and the language lacks them.

5.5 Conclusions

In this chapter a particular parametric approach consistent with the learnability conditions under consideration in chapter 3 has been proposed: a mechanism of morphophonological data analysis, the Chunking Procedure, active during language acquisition.

We have investigated to what extent the value of perceptually available features obtained by the Chunking Procedure, specifically how they are phonologically packed [±bound] and what their morphological nature is [±synthetic], eases the acquisition of certain abstract linguistic properties. In other words, we have explored how the analysis obtained by the Chunking

Procedure could be used to bootstrap the presence of higher order syntactic patterns of target languages.

In particular, we have proposed three bootstrapping mechanisms that are triggered by the analysis made by the Chunking Procedure. First, it has been proposed that the learner can infer, from the analysis of certain morphs, syntactic properties traditionally assigned to Polysynthetic languages, and we have shown how this mechanism can also account for mixed patterns and "polysynthetic-like" structures. Second, it has been argued that the learner can bootstrap radical pro-drop patterns in his target language from the analyses attained by the Chunking Procedure in examining certain pronominal affixes. Third, we have proposed that the learner can infer the possibility of having some resultatives constructions from the analysis obtained regarding the morph instantiating *path*.

An interesting contribution of such a proposal is to provide a parametric definition of some mechanisms of the morphophonological analyser taking into consideration both learnability considerations as well as some well-known patterns of linguistic variation. Indeed, whereas we capitalise on the former, the latter becomes derivative. By using this methodology, linguistic variation can be explored with the same means as those arguably at work in acquisition when the learner analyses his linguistic input. Actually morphosyntactic variation is argued to be constrained by mechanisms of data analysis active during the process of language acquisition, as we have been proposing all along in this thesis.

It is worth remarking that this specific parametric approach does not formulate parameters that compress multiple syntactic properties, thereby differing from the standard conception of parameter (see chapter 3 for an overview). Instead we have directly defined mechanisms of data analysis, which must independently be at work, using two atomic morphophonological properties of heads: their [±boundedness] and their [±syntheticity]. The resulting analysis feeds certain bootstrapping mechanisms, used by the LAD to infer the clusters of syntactic properties which traditional parameters compress.

Furthermore, provided that procedures of data analysis are considered to be elements of Third Factor (Chomsky 2005), this proposal leads us to the appealing conclusion that by defining data analysers in a parametric fashion, linguistic variation could be embodied in certain Third Factor mechanisms. Combining this idea with Chomky's Uniformity Hypothesis, the proposal put forward in this chapter emphasises the invariant nature of the syntactic and the semantic components: linguistic variation is a matter of how languages pack their morphophonological content.

A novelty of this chapter with respect Fasanella and Fortuny (2012) is that some inter-theoretical explanations of why the analyses obtained by the Chunking Procedure correlate with certain bootstrapping mechanisms have been included. As has been noted previously, the present proposal is not contingent on these particular explanations, but it may be of interest to investigate the theoretical reasons for which perceptible properties cross-linguistically correlate with morphosyntactic patterns.

In this chapter we have not explored the relationship between the analyses obtained by the learner using the Chunking Procedure and patterns of morphosyntactic change. Nonetheless, it may be argued that learners' analyses regarding the [±boundedness] and the [±syntheticity] of morphs are responsible for common attested patterns of diachronic morphosyntax. Actually, some of the most well-known morphosyntactic changes observed in natural languages seem to involve the [±boundedness] and the [±syntheticity] variables; a fact which is completely consistent with the assumption that segmenting the string and assigning semantic content to morphs are the two fundamental tasks that have to be carried out by the learner when acquiring morphology, as we already pointed out in chapter 2.

Indeed it is very easy to find examples of these types of changes in the literature. Just to consider a few cases, it is enough to take a look at Campbell (1998)'s list of recurrent diachronic changes in world languages, which very often involve a novel segmentation of linguistic pieces and/or a novel distribution of morphemes with respect to morphs: it is very frequently observed that case suffixes derive from postpositions, as the Estonian commitative suffix /-ka/ derived from the postposition 'with' *kanssa*; that the future tense

affix marker also derive from an independent verb, sometimes the verb 'want', as the English 'will'; that bound auxiliary verbs derive from main verbs; that durative/habitual morphs derive from full verbs, such as 'stay'; that the dative case marker sometimes derive from an independent verb, frequently 'give'; or that definite articles derive from demonstratives, as in the case of most Romance languages.

The mechanism of data analysis explored in this chapter, the Chunking Procedure, can be viewed as an instance of a common move in the history of Generative Grammar. Actually, the type of parameters commonly postulated in the literature to cover the considered linguistic facts have a status similar to certain abandoned artifacts, such as constructions, rewriting rules or transformation rules, inasmuch as they describe properties of grammatical byproducts. Nevertheless, in order to understand the nature of UG it is not enough to determine the characteristics of constructs: we must unearth the fundamental components that are behind them.

6

General discussion

In this thesis we have argued that mechanisms in charge of language learning, independently postulated on learnability grounds, shape the format of variation among natural languages, both synchronically and diachronically.

We have offered theoretical arguments in favour of such an enterprise and we have proposed two devices that can arguably be used by the LAD. We have examined how these devices, apart from being at work in acquisition, can be responsible for some patterns of variation among natural languages. We have tried to show, then, how Greenberg's problem can be successfully approached as a side effect of Plato's problem.

In chapter 2 the aim of this dissertation was presented, as well as the basic definitions, assumptions and methodology. Since the beginning, we have emphasised the importance of the independent postulation of learning devices based on learnability grounds before arguing that they are responsible for some concrete pattern of variation.

Chapter 3 was devoted to reviewing proposals that focused either on Greenberg's problem or on Plato's problem. We have shown that (macro-/micro-) parametric models neglect some proposed learnability conditions. Regarding learnability models, it has been argued that they cannot be considered plausible models for the acquisition of natural languages since the idealised languages they can operate with are excessively simplified to handle very spread patterns found across natural languages. After this review, we sketched the basics for a plausible approach of the language acquisition

task that not only follows learnability considerations, but is also able to account for linguistic patterns of variation. Building on Chomsky (1957)'s tripartite classification, we have proposed to conceive the LAD as a discovery procedure, and we have made explicit some of the elements that this procedure would incorporate: mechanisms of data analysis, principles of efficient computation and learnability constraints (figure 3.7 repeated here as figure 6.1):

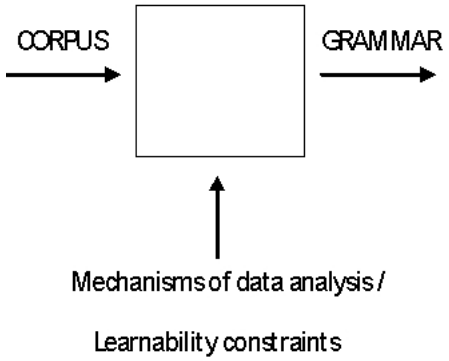


Figure 6.1: The modified discovery procedure

In chapter 4 a learning restriction derived from simplicity considerations was proposed to be used by the learner when acquiring inflectional systems, the Minimise Paradigms Constraint (MPC). The use of the MPC by the LAD allows the learner the acquisition of simpler representations as well as a simpler analysis of the received input, where simplicity is measured following Bane (2008). We have also argued that successive analyses of learners using the MPC have predictable results in the I-languages finally acquired. Specifically, it has been argued that fusional and agglutinative patterns in the morphology of natural languages can be the result of the effects of the MPC. In order to illustrate the first scenario, namely, the effects of the MPC in deriving fusional patterns, we have been concerned with diachronic changes from Latin to some Romance verbal inflectional systems. Regarding the illustration of how the MPC predicts agglutinative patterns, we have focused on the Chichewa verbal system. Apart from these two case studies, we have

pointed out to two other inflectional systems that can be considered for further research on how the MPC works: Hua and Turkish verbal complexes.

Chapter 5 was dedicated to extending the proposal in Fasanella and Fortuny (2012), where a morphophonological mechanism of data analysis, the Chunking Procedure, was argued to be used by the learner when acquiring morphology, and whose results can be used as triggers for the bootstrapping of certain morphosyntactic properties found among natural languages. It has been proposed that some of the properties traditionally associated to a) the so-called Polysynthesis Parameter, b) radical pro-drop, and c) satellite-framed languages, can be linked to the morphophonological results obtained using the Chunking Procedure. Accordingly, the acquisition of high-order properties is linked to the acquisition of morphophonological perceptible features of languages.

As pointed out at the beginning of chapter 2, the whole enterprise that this project stands for cannot be (completely) developed in a single dissertation. However, we conceive of the present study as a contribution, both theoretical and empirical, towards that specific general goal. Actually, as argued in chapter 3, parametric theory within the Government and Binding period and also afterwards was designed to handle the question of language acquisition and linguistic variation at once, although most of the concrete proposals were just concerned with one of the two problems. Given that, the present study can be understood as a tentative effort to investigate this objective following the original rationale that once was at the first line of the research agenda in Generative linguistics. More precisely, what we have tried to show is that in order to meet both descriptive and explanatory adequacy, the former must be considered as derivative from the later.

In this respect, one of the novelties of the proposal we have argued for is that we can also go "beyond explanatory adequacy" by adopting this concrete framework. The format and degree of linguistic variation, that is, the concrete patterns adopted by natural languages in instantiating a specific final state of the Language Faculty, come from a language-internal area: the domain of acquisition. This is, perhaps, the major general contribution of the present dissertation.

Regarding some concrete contributions, we may emphasise the incorporation of clear definitions of the notions 'morph' and 'morpheme', the proposal of three learnability conditions, the development of two particular instances of Third Factor mechanisms (Chomsky 2005), the MPC and the Chunking Procedure, and the analysis of very well-known patterns of the Latin and Romance verbal systems under a different perspective.

Let us now point out some aspects that have been explicitly or implicitly left for further research.

In chapter 3 we conclude with the imperative of arguing for a modified discovery procedure where abstract linguistic aspects were acquired from more perceptual features in the input. In the remaining chapters we have been mainly concerned with morphology, and how some morphophonological aspects can be the trigger for the acquisition of less accessible morphosyntactic properties. Hopefully in the near future this methodology could be fruitfully used in other domains, as for example, in grounding the acquisition of high-order linguistic properties from salient prosodic cues.

We consider that one of the most appealing ideas of the present thesis was put forward in chapter 4, when considering distributional properties as relevant cues for the acquisition of semantic aspects of morphs; more concretely, for the acquisition of the ±agglutinative nature of morphs. It would be interesting, then, to continue investigating this line of research: first, more data should be considered with respect to this in order to show that the MPC makes predictions not only in verbal morphology, but also in the nominal domain; second, experimental work should shed light on how the learner obtains distributional information regarding the the MPC and how this information is used to trigger the acquisition of the semantics of morphs.

As for the effects of the MPC, another aspect that may be of interest for further research is related to the observation that "postpositions frequently become attached to roots and lose their independent status, becoming case suffixes; however, case suffixes hardly ever become independent postpositions" (Campbell 1998, 303). It would be worth investigating if the MPC or another simplifying device of a similar sort could be responsible for such a robust diachronic observation: is this related to the assumption that the

learner prefers to hypothesise less cuts in the input string and takes advantage of syntagmatic relationships in order to postulate as many [+bound] morphs as possible?

In chapter 5 we provide synchronic data to support our claims, though in the conclusions some diachronic cases were briefly pointed out to be considered for further work in possible extensions of the Chunking Procedure. We think that it would be very interesting to follow this line and study the relationship of this concrete mechanism of morphophonological data analysis with some patterns of language change. Indeed, diachronic change can provide strong arguments in favour of such a mechanism. Let us make this point explicit in some detail. We have argued that morphophonological properties of morphs, their [±boundedness] and their [±syntheticity], correlate with syntactic patterns, and we have illustrated this statement with three case studies. If this hypothesis is in fact true, and these properties of morphs correlate with syntactic patterns, we would expect that, when a relevant morph suffered a modification in its [±boundedness] and/or its [±syntheticity], the concrete syntactic patterns associated with the former value of this morph, also changed in the direction implicitly predicted by the postulated bootstrapping mechanisms.

Fortunately, it is possible to illustrate this hypothesis with one of the three case studies considered in chapter 5: the morphological realisation of the morph encoding *path* and their related syntactic constructions. In a former stage of Latin, *path* was realised by [-synthetic] morphs, and this is why Latin is considered a (weak) satellite-framed language. In being a satellite-framed language, Latin shows the set of syntactic constructions associated to these types of languages, as argued and illustrated by Acedo-Matellán (2010). However, in subsequent instances of the language, in Romance, the morph encoding *path* become fusionalised with the verb, that is, it stopped being realised in a satellite of the verb. As predicted, syntactic constructions associated to the independent lexicalisation of *path* were not available in the language any more. Therefore, this specific association of morphological properties with syntactic properties is corroborated by diachronic data. It would be desirable, then, to apply this methodology to the other case studies

considered in chapter 5, investigating the following research questions: a) can it be diachronically shown that when a previously [-bound] morph becomes [+bound] its associated maximal projection becomes free in position and not obligatory?; b) can it be diachronically proved that when a previously [+synthetic] (fusional) morph encoding *case* becomes [+bound] [-synthetic] (agglutinative), the language allows the omission of subjects, objects and possessors?

Also related to the Chunking Procedure is the observation first made in chapter 2 that, for the acquisition of morphology, the learner has to isolate meaningful pieces and map consistent meanings onto them. It would be interesting to investigate how this rather standard assumption and the dynamics of the Chunking Procedure relate to diachronic changes where morphs are lost. As Nurse (2007) observes "if a language loses the segmental expression of derivational and aspectual categories, in principle it faces a choice: it can also lose the categories, or it can express them some other way. In practice, this seems to be hardly a choice as all the languages I have looked at that have lost extensions (...), keep the categories and express them some other way" (Nurse 2007, 240). This means that morphemes ('categories' in Nurse's terminology) are pervasive and change from morph_1 to morph_2 when morph_1 is lost.[1] This cross-linguistic observation is related to the role of the Chunking Procedure and reminds us about some diachronic cyclical changes where a particular morpheme is expressed by an isolated morph ([-bound]), then by an affixated one ([+bound] [-synthetic]) and finally fusioned to another piece ([+synthetic]), as we have observed in some detail in relation to the change from Latin to Romance regarding the expression of *path*. Further research should show how cyclical changes are related with the two basic tasks in the acquisition of morphology, namely, segmenting and mapping meaning, as specified by the two variables contained in the Chunking Procedure, the [±boundedness] and the [±syntheticity] of morphs.

[1] Another alternative, not considered here but possible, is that a morpheme originally expressed by a morph_1 that was lost becomes to be realised suprasegmentally, and not by a morph_2.

The arguments and empirical observations compiled in this thesis further support an ideal that commonly guides the history of Generative Grammar, although it is rarely entertained in most investigations: by seriously approaching the logical problem of language acquisition, or Plato's problem, fundamental aspects of the Faculty of Language and of natural languages can be discovered.

References

Acedo-Matellán, V. 2010, *Argument Structure and the Syntax-Morphology Interface. A Case Study in Latin and Other Languages*, PhD thesis, Universitat de Barcelona.

Adams, M. 1987, From Old French to the Theory of Pro-Drop, *Natural Language & Linguistic Theory* **5**(1), 1–32.

Alkire, T. and Rosen, C. 2010, *Romance Languages: a Historical Introduction*, Cambridge University Press.

Allen, B., Gardiner, D. and Frantz, D. 1984, Noun Incorporation in Southern Tiwa, *International Journal of American Linguistics* **50**(3), 292–311.

Baker, M. 1988, *Incorporation: a Theory of Grammatical Function Changing*, University of Chicago Press.

Baker, M. 1996, *The Polysynthesis Parameter*, Oxford University Press.

Baker, M. 2001, *The Atoms of Language.*, Basic books..

Baker, M. 2005, The Innate Endowment for Language: Underspecified or Overspecified? Ms., University of Rutgers.

Baker, M. 2008, The Macroparameter in a Microparametric World, *in* T. Biberauer (ed.), *The Limits of Syntactic Variation*, John Benjamins, 351–374.

Bane, M. 2008, Quantifying and Measuring Morphological Complexity, *Proceedings of the 26th West Coast Conference on Formal Linguistics*, 69–76.

Bauer, L. 2004, *A Glossary of Morphology*, Edinburgh University Press.

REFERENCES

Berwick, R. 1985, *The Acquisition of Syntactic Knowledge*, MIT press.

Berwick, R., Chomsky, N. and Piattelli-Palmarini, M. 2013, Poverty of the Stimulus Stands: Why Recent Chanllenges Fail, *in* R. Berwick and M. Piattelli-Palmarini (eds.), *Rich Languages from Poor Inputs*, Oxford University Press.

Biberauer, T., Roberts, I. and Sheehan, M. 2013, No-Choice Parameters and the Limits of Syntactic Variation, *Proceedings of the 31st West Coast Conference on Formal Linguistics*.

Bickel, B. and Nichols, J. 2005, Inflectional Synthesis of the Verb, *The World Atlas of Language Structures*, 94–97.

Blevins, J. and Blevins, J. 2009, *Analogy in Grammar: Form and Acquisition*, Oxford University Press.

Boeckx, C. 2011, Approaching Parameters from Below, *in* C. Boeckx and A.-M. D. Sciullo (eds.), *The Biolinguistic Enterprise: New Perspectives on the Evolution and Nature of the Human Language Faculty*, Oxford University Press.

Brown, R. 1973, *A First Language: The Early Stages*, Harvard University Press.

Caha, P. 2009, *The Nanosyntax of Case. PhD thesis*, University of Tromsø.

Campbell, L. 1998, *Historical Linguistics: an Introduction*, MIT Press.

Carstairs-McCarthy, A. 2001, Paradigmatic Structure: Inflectional Paradigms and Morphological Classes, *in* A. Spencer and A. Zwicky (eds.), *The Handbook of Morphology*, Oxford University Press.

Chomsky, N. 1957, *Syntactic Structures*, Mouton.

Chomsky, N. 1959, A Review of B. D. Skinner's *Verbal Behavior*, *Language* **35**(1), 26–58.

Chomsky, N. 1965, *Aspects of the Theory of Syntax*, MIT press.

Chomsky, N. 1977, On *Wh*-Movement, *in* P. Culicover, T. Wasow and A. Akmajian (eds.), *Formal Syntax*, Academic Press, 71–133.

Chomsky, N. 1981, *Lectures on the Theory of Government and Binding*, Foris.

Chomsky, N. 1993, A Minimalist Program for Linguistic Theory, *in* K. Hale and S. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, MIT Press, 1–52.

Chomsky, N. 1995, *The Minimalist Program*, MIT Press.

Chomsky, N. 2001, Derivation by Phase. *in* M. Kenstowicz (ed.), *Ken Hale: a Life in Language*, MIT Press, 1–52.

Chomsky, N. 2001b, Beyond Explanatory Adequacy, *MIT Occasional Papers in Linguistics* (20), MIT Press, 1–28.

Chomsky, N. 2002, Reedition of *Syntactic Structures*, Mouton.

Chomsky, N. 2005, Three Factors in Language Design, *Linguistic Inquiry* **36**(1), 1–22.

Christophe, A., Dupoux, E., Bertoncini, J. and Mehler, J. 1994, Do Infants Perceive Word Boundaries? An Empirical Study of the Bootstrapping of Lexical Acquisition, *Acoustical Society of America* **95**, 1570–1570.

Cinque, G. 1999, *Adverbs and Functional Heads: A Cross-Linguistic Perspective*, Oxford University Press.

Clark, E. 2001, Morphology in Language Acquisition, *in* A. Spencer and A. Zwicky (eds.), *The Handbook of Morphology*, Oxford University Press, 374–389.

Clark, R. 1992, The Selection of Sntactic Knowledge, *Language Acquisition* **2**(2), 83–149.

Clark, R. and Roberts, I. 1993, A Computational Model of Language Learnability and Language Change, *Linguistic Inquiry* **24**(2), 299–345.

Culicover, P. and Wexler, K. 1977, Some Syntactic Implications of a Theory of Language Learnability, *in* P. Culicover and A. Akmajian (eds.), *Formal Syntax*, Academic Press.

Culbertson, J. 2010, *Learning Biases, Regularization, and the Emergence of Typological Universals in Syntax*, PhD thesis, Johns Hopkins University.

De Rijk, R. 2007, *Standard Basque: A Progressive Grammar*, MIT Press.

Dresher, B. 1999, Charting the Learning Path: Cues to Parameter Setting, *Linguistic Inquiry* **30**(1), 27–67.

Dresher, B. and Kaye, J. 1990, A Computational Learning Model for Metrical Phonology, *Cognition* **34**(2), 137–195.

Embick, D. and Noyer, R. 1999, Locality in Post-Syntactic Operations, *MIT Working Papers in Linguistics* **34**, MIT Press, 265–317.

Embick, D. and Noyer, R. 2001, Movement Operations after Syntax, *Linguistic Inquiry* **32**(4), 555–595.

Fasanella, A. 2011, Els problemes de Plató, de Darwin i de Greenberg, *Els Marges* **94**, 18–36.

Fasanella, A. and Fortuny, J. 2012, Deriving Linguistic Variation from Learnability Conditions: the Chunking Procedure. Ms., Universitat Autònoma de Barcelona.

Foley, W. 1991, *The Yimas Language of New Guinea*, Stanford University Press.

Fortuny, J. 2008, *The Emergence of Order in Syntax*, John Benjamins.

Fowler, C. 1977, *Timing Control in Speech Production*, Indiana University.

Gervain, J. and Mehler, J. 2010, Speech Perception and Language Acquisition in the First Year of Life, *Annual Review of Psychology* **61**, 191–218.

Gianollo, C., Guardiano, C. and Longobardi, G. 2008, Three Fundamental Issues in Parametric Linguistics, *in* T. Biberauer (ed.), *The Limits of Syntactic Variation*, John Benjamins, 109–142.

Gibson, E. and Wexler, K. 1994, Triggers, *Linguistic Inquiry* **25**(3), 407–454.

Gilligan, G. 1987, *A Cross-Linguistic Approach to the Pro-Drop Parameter*, PhD thesis, University of Southern California.

Gleitman, L. and Fisher, C. 2005, Universal Aspects of Word Learning, *in* J. McGilvray (ed.), *The Cambridge Companion to Chomsky*, Cambridge University Press.

Gold, M. 1967, Language Identification in the Limit, *Information and Control* **10**(5), 447–474.

Gopnik, A. 1996, The theory theory as an alternative to the innateness hypothesis, *in* L. Antony and N. Hornstein (eds.), *Chomsky and his Critics*, Blackwell, 238–253.

Güldemann, T. 2003, Grammaticalization, *The Bantu Languages*, Routledge Language Family Series, 182–194.

Haiman, J. 1980, *Hua, a Papuan Language of the Eastern Highlands of New Guinea*, John Benjamins.

Haiman, J. 2001, Hua (Papuan), *in* A. Spencer and A. Zwicky (eds.), *The Handbook of Morphology*, Oxford University Press, 539–562.

Hale, K. and Keyser, S. 1998, The Basic Elements of Argument Structure, *MIT Working Papers in Linguistics* **32**, 73–118.

Halle, M. and Marantz, A. 1993, Distributed Morphology and the Pieces of Inflection, *in* K. Hale and J. Keyser (eds.), *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, MIT Press.

Halle, M. and Marantz, A. 1994, Some Key Features of Distributed Morphology, *in* A. Carnie and H. Harley (eds.), *MIT Working Papers in Linguistics* **21**, MIT Press.

Hamburger, H. and Wexler, K. 1975, A Mathematical Theory of Learning Transformational Grammar, *Journal of Mathematical Psychology* **12**(2), 137–177.

Harris, Z. 1968, *Mathematical Structures of Language*, Wiley.

Harris, Z. 1970, Distributional Structure, *Papers in Structural and Transformational Linguistics*, 775–794.

Hauser, M., Chomsky, N. and Fitch, W. 2002, The Faculty of Language: What Is It, Who Has It, and How Did It Evolve?, *Science* **298**(5598), 1569–1579.

Herman, J. 2000, *Vulgar Latin*, Penn State University Press.

Hoji, H. 1998, Null Object and Sloppy Identity in Japanese, *Linguistic Inquiry* **29**(1), 127–152.

Holmberg, A. and Roberts, I. 2010, *Macroparameters and the Minimalist Program*, Talk given at Linguistic Variation in the Minimalist Framework Workshop, Universitat Autònoma de Barcelona.

Huang, J. 1984, On the Distribution and Reference of Empty Pronouns, *Linguistic Inquiry*, 531–574.

Hyams, N. 1983, *The Acquisition of Parameterized Grammars*, PhD thesis, City University of New York.

Jelinek, E. 1984, Empty Categories, Case, and Configurationality, *Natural Language & Linguistic Theory* **2**(1), 39–76.

Joseph, B. 2001, Diachronic Morphology, *in* A. Spencer and A. Zwicky (eds.), *The Handbook of Morphology*, Oxford University Press, 351–373.

Kato, M. 1999, Strong and Weak Pronominals in the Null Subject Parameter, *Probus* **11**(1), 1–37.

Kayne, R. 2000, *Parameters and Universals*, Oxford University Press.

Kayne, R. 2005a, On Parameters and on Principles of Pronunciation. Ms., New York University.

Kayne, R. 2005b, Some Notes on Comparative Syntax, with Special Reference to English and French, *The Oxford Handbook of Comparative Syntax*, Oxford University Press, 3–69.

Kemps, R., Wurm, L., Ernestus, M., Schreuder, R. and Baayen, H. 2005, Prosodic Cues for Morphological Complexity in Dutch and English, *Language and Cognitive Processes* **20**(1-2), 43–73.

Kiparsky, P. 1973, "Elsewhere" in Phonology, *in* S. Anderson and P. Kiparsky (eds.), *A Festschrift for Morris Hale*, Holt, Rinehart and Winston.

Kolmogorov, A. 1965, Three Approaches to the Quantitative Definition of Information, *Problems of Information Transmission* **1**(1), 1–7.

Kornfilt, J. 1997, *Turkish*, Routledge.

Lakoff, G. and Johnson, M. 1980, *Metaphors We Live By*, University of Chicago Press.

Ledgeway, A. 2012, *From Latin to Romance: Morphosyntactic Typology and Change*, Oxford University Press.

Legate, J. 2002, *Warlpiri: Theoretical Implications*, PhD thesis, MIT.

Lehiste, I. 1970, *Suprasegmentals*, MIT press.

Lightfoot, D. 1979, Principles of Diachronic Syntax, *Cambridge Studies in Linguistics* **23**.

Lindblom, B. and Rapp, K. 1973, *Some Temporal Regularities of Spoken Swedish*, University of Stockholm.

Longobardi, G. 1994, Reference and Proper Names: a Theory of N-Movement in Syntax and Logical Form, *Linguistic Inquiry*,**25**(4), 609–665.

Longobardi, G. 2013, *How Universal is DP?*, Talk at Going Romance 2012, University of Amsterdam.

Marcus, G., Vijayan, S., Bandi Rao, S. and Vishton, P. 1999, Rule Learning by Seven-Month-Old Infants, *Science* **283**(5398), 77–80.

Mascaró, J. 1986, *Morfologia*, Vol. 5, Enciclopèdia Catalana.

Mchombo, S. 2001, Chichewa (Bantu), *in* A. Spencer and A. Zwicky (eds.), *The Handbook of Morphology*, Oxford University Press.

McIntyre, A. 2004, Event Paths, Conflation, Argument Structure, and vP Shells, *Linguistics*, **42**(3), 523–572.

McWhorter, J. 2001, The Worlds Simplest Grammars Are Creole Grammars, *Linguistic Typology* **5**(2,3), 125–166.

Meeussen, A. 1967, Bantu Grammatical Reconstructions, *Africana Linguistica* **3**, 79–121.

Mendívil-Giró, J. L.: 2012, The Myth of Language Diversity. Language Unity and Diversity from a Biolinguistic Point of View, *in* M. C. Horno-Cheliz, C. Boeckx and J. L. Mendívil (eds.), *Language, from a Biological Point of View*, Cambridge Scholars Publishing.

Neeleman, A. and Szendrői, K. 2007, Radical Pro Drop and the Morphology of Pronouns, *Linguistic Inquiry* **38**(4), 671–714.

Newmeyer, F. 2005, *Possible and Probable Languages: A Generative Perspective on Linguistic Typology*, Oxford University Press.

Nicolis, M. 2008, The Null Subject Parameter and Correlating Properties: the Case of Creole Languages, *in* T. Biberauer (ed.), *The Limits of Syntactic Variation*, John Benjamins, 271–294.

Niyogi, P. and Berwick, R. 1996, A Language Learning Model for Finite Parameter Spaces, *Cognition* **61**(1), 161–193.

Nooteboom, S. 1972, *Production and Perception of Vowel Duration: a Study of Durational Properties of Vowels in Dutch*, Philips Research Laboratories.

Nowak, M. and Krakauer, D. 1999, The Evolution of Language, *Proceedings of the National Academy of Sciences* **96**(14), 8028–8033.

Nurse, D.: 2007, Did the Proto-Bantu Verb Have a Synthetic or an Analytic Structure?, *Bantu in Bloomsbury: Special Issue on Bantu linguistics*, SOAS Working Papers in Linguistics, **15**, 239–256.

Pinker, S. 1979, Formal Models of Language Learning, *Cognition* **7**(3), 217–283.

Pinker, S. 1982, A Theory of the Acquisition of Lexical Interpretive Grammars, *in* J. Bresnan (ed.), *The Mental Representation of Grammatical Relations*, MIT Press.

Pinker, S. 1984, *Language Learnability and Language Development*, Harvard University Press.

Real-Puigdollers, C. 2013, *Lexicalization by Phase: the Role of Prepositions in Argument Structure and its Cross-Linguistic variation*, PhD thesis, Universitat de Autònoma de Barcelona.

Redington, M. and Chater, N. 1998, Connectionist and Statistical Approaches to Language Acquisition: a Distributional Perspective, *Language and Cognitive Processes* **13**(2-3), 129–191.

Redington, M., Crater, N. and Finch, S. 1998, Distributional Information: a Powerful Cue for Acquiring Syntactic Categories, *Cognitive Science* **22**(4), 425–469.

Rizzi, L. 1982, *Issues in Italian Syntax*, Foris.

Rizzi, L. 1986, Null Objects in Italian and the Theory of pro, *Linguistic Inquiry* **17**(3), 501–557.

Roberts, I. 1993, *Verbs and Diachronic Syntax: a Comparative History of English and French*, Kluwer.

Roberts, I. 2007, *Diachronic Syntax*, Oxford University Press.

Roberts, I. 2010, *Agreement and Head Movement*, The MIT Press.

Roberts, I. and Holmberg, A. 2010, Introduction: Parameters in Minimalist Theory, *in* T. Biberauer. et al. (eds.), *Parametric Variation. Null Subjects in Minimalist Theory*, Cambridge University Press.

Roberts, I. and Roussou, A. 2003, *Syntactic Change: A Minimalist Approach to Grammaticalization*, Cambridge University Press.

Rubio, C. and Fasanella, A. 2014, *Radical Pro-Drop, Lack of Verbal Agreement and Agglutinative Morphology. A Unified Account* Talk given at ConSOLE XXII, University of Lisbon.

Saffran, J., Aslin, R. and Newport, E. 1996, Statistical Learning by 8-Month-Old Infants, *Science* **274**(5294), 1926.

Sahlgren, M. 2008, The Distributional Hypothesis. From context to Meaning: Distributional Models of the Lexicon in Linguistics and Cognitive Science, *Rivista di Linguistica* **20**(1), 33–54.

Saito, M. 2007, Notes on East Asian Argument Ellipsis, *Language Research* **43**(2), 203–227.

Sakas, W. and Fodor, J. 2001, The Structural Triggers Learner, *in* S. Bertolo (ed.), *Language Acquisition and Learnability*, Cambridge University Press.

Saltarelli, M., Azkarate, M., Farwell, D., de Urbina J. and Onedera, L. 1988, *Basque*, Croom Helm London.

Sansavini, A., Bertoncini, J. and Giovanelli, G. 1997, Newborns Discriminate the Rhythm of Multisyllabic Stressed Words., *Developmental Psychology* **33**(1), 3.

Sapir, E. 1921, *Language*, Harcourt, Brace and World.

Sezer, E. 2001, Finite Inflection in Turkish, *in* E. Taylan (ed.), *The Verb in Turkish*, John Benjamins.

Shosted, R. 2006, Correlating Complexity: A Typological Approach, *Linguistic Typology* **10**(1), 1–40.

Slobin, D. 1984, *The Crosslinguistic Study of Language Acquisition*, Erlbaum Associates.

Snyder, W. 2001, On the Nature of Syntactic Variation: Evidence from Complex Predicates and Complex Word-Formation, *Language* **77**(2), 324–342.

Snyder, W. 2012, Parameter Theory and Motion Predicates, *in* V. Demonte and L. McNally (eds.), *Telicity, Change, and State*, Oxford University Press.

Son, M. 2007, *Directionality and Resultativity: The Cross-Linguistic Correlation Revisited*, Nordlyd, TromsøWorking Papers on Language and Linguistics.

Speas, P. 1994, Null Arguments in a Theory of Economy of Projection, *University of Massachusetts Occasional Papers in Linguistics* .

Talmy, L. 1985, Lexicalization Patterns: Semantic Structure in Lexical Forms, *Language Typology and Syntactic Description* **3**, 57–149.

Talmy, L. 2000, *Toward a Cognitive Semantics, Vol. 2: Typology and Process in Concept Structuring.*, MIT Press.

Trask, L. 2002, Book review: The Atoms of Language: The Mind's Hidden Rules of Grammar by Mark C. Baker, *The Human Nature Review* **2**, 77–81.

Väänänen, V. 1982, *Introducción al latín vulgar*, Gredos.

Valentí i Fiol, E. 1995, *Gramàtica llatina: morfologia i nocions de sintaxi*, Curial.

Valian, V. 1991, Syntactic Subjects in the Early Speech of American and Italian Children, *Cognition* **40**(1), 21–81.

Wexler, K. and Hamburger, H. 1975, Learning-Theoretic Foundations of Linguistic Universals, *Theoretical Linguistics* (2), 215–253.

Yang, C. 2002, *Knowledge and Learning in Natural Language*, Oxford University Press.

Yang, C. 2004, Universal grammar, Statistics or Both?, *Trends in Cognitive Sciences* **8**(10), 451–456.

Yang, C. 2005, On productivity, *Linguistic Variation Yearbook* **5**(1), 265–302.

Yang, C. 2006, *The Infinite Gift: How Children Learn and Unlearn the Languages of the World*, Scribner Book Company.

Yang, C. 2010, Three Factors in Language Variation, *Lingua* **120**(5), 1160–1177.

Zubizarreta, M. and Oh, E. 2007, *On the syntactic composition of manner and motion*, MIT Press.

Declaration

This PhD Thesis has been supported by the the grants *Formació de personal investigador* (FI) (2010FI-B 00785) by the Generalitat de Catalunya from February to September 2010, and *Formación de Personal Universitario* (FPU) (AP2009-1396) by Gobierno de España from October 2010 to January 2014.

This PhD Thesis has also been supported by the projects *Microvariación: rasgos sintácticos y realización morfofonológica* HUM 2006-13295-CO2-01/FILO (principal investigator: Carme Picallo), *La estructura de la periferia infraoracional* FFI 2011-29440-C03-03 (principal investigator: Anna Bartra), and *Grup d'investigació Lingüística Teòrica* (1370)-2009-SGR1079 (principal investigator: Gemma Rigau).