

REFERENCE ONLY



2809586038

## UNIVERSITY OF LONDON THESIS

Degree phd

Year 2007

Name of Author MARCO

TAMBURELLI

### COPYRIGHT

This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting the thesis must read and abide by the Copyright Declaration below.

### COPYRIGHT DECLARATION

I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

### LOAN

Theses may not be lent to individuals, but the University Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: The Theses Section, University of London Library, Senate House, Malet Street, London WC1E 7HU.

### REPRODUCTION

University of London theses may not be reproduced without explicit written permission from the University of London Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

- A. Before 1962. Permission granted only upon the prior written consent of the author. (The University Library will provide addresses where possible).
- B. 1962 - 1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.
- C. 1975 - 1988. Most theses may be copied upon completion of a Copyright Declaration.
- D. 1989 onwards. Most theses may be copied.

***This thesis comes within category D.***

☐

This copy has been deposited in the Library of

UCL

☐

This copy has been deposited in the University of London Library, Senate House, Malet Street, London WC1E 7HU.



# The Role of Lexical Acquisition in Simultaneous Bilingualism

Thesis submitted to the University of London in partial fulfilment  
of the requirements for the degree of  
Doctor of Philosophy

**Marco Tamburelli**

Department of Phonetics and Linguistics  
University College London

2007

UMI Number: U593460

All rights reserved

INFORMATION TO ALL USERS

The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript and there are missing pages, these will be noted. Also, if material had to be removed, a note will indicate the deletion.



UMI U593460

Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.  
Microform Edition © ProQuest LLC.

All rights reserved. This work is protected against  
unauthorized copying under Title 17, United States Code.



ProQuest LLC  
789 East Eisenhower Parkway  
P.O. Box 1346  
Ann Arbor, MI 48106-1346



I, Marco Tamburelli, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

## **Abstract**

This thesis addresses a central issue within the field of Bilingual First Language Acquisition (BFLA) by exploring the extent to which the two languages can affect each other during development. The aim of this work is twofold. Firstly, it proposes a particular formalisation of the acquisition process which is based on a set of assumptions that are mostly drawn from standard linguistic theory. It then argues for a theory of the acquisition of lexical properties that is based on the interaction between two higher level systems. The first of these is a system dedicated to organising the developing lexicon into paradigms while the second is an informationally monotonic updating system whose role is to add newly acquired lexical information to those items that are not yet fully developed.

It is then argued that this model can accommodate transfer effects as an inevitable consequence of BFLA. Given that the lexicon of a bilingual child is larger than that of monolinguals, the updating mechanism has a wider field of application and therefore - besides over-generalisation - transfer effects will also obtain. An important consequence of this claim is that the only difference between a monolingual and a bilingual child lies in the domain within which the updating mechanism applies. The fact that language production in bilingual children differs from that of monolinguals does not force the postulation of special bilingual strategies but can be accounted for by appealing to the very two aspects that monolinguals and bilinguals do not have in common, namely the input and the number of developing lexical sets.

A substantial part of the thesis is dedicated to evaluating the empirical coverage of this model. This involves discussion of data from case-studies as well as experimental work both old and new.

# Table of Contents

---

<b>Acknowledgements</b>	<b>9</b>
<b>INTRODUCTION</b>	<b>10</b>
<b>CHAPTER 1: On the Acquisition of Lexical Properties</b>	
<b>1. Introduction</b>	<b>14</b>
<b>2. The Acquisition Process</b>	<b>14</b>
2.1 Preamble	14
2.2 Lexical Properties and Their Value	17
2.3 Language Variation	21
2.4 Acquiring Language-Specific Idiosyncrasies	23
2.4.1 Introduction	23
2.4.2 The Notion of Paradigm	25
2.4.3 Towards a Paradigmatic Acquisition	27
2.4.4 The Role of Paradigm Formation	31
2.4.5 UG-Constrained Updating	35
<b>3. Updating and Overgeneralisation</b>	<b>38</b>
3.1 Introduction	38
3.2 Acquisition and Innateness	38
3.2.1 The Initial State	38
3.3 Mapping Principles	40
3.3.1 Variable Binding and Syntax/Semantics Mapping	41
3.4 The Transitive/Unaccusative Alternation	43
3.5 The Nature of Overgeneralisation	45
3.6 The Simplex/Complex Alternation	47
3.7 Alternative Analysis: Morphological Affixation	50
<b>4. Retreating from a Non-target Lexicon</b>	<b>51</b>
4.1 Introduction	51
4.2 Deleting Information	52

<b>5. Interim Summary</b>	<b>55</b>
<b>6. Bilingual Acquisition</b>	<b>56</b>
6.1 How Can a System Deal with Two Languages?	56
6.1.1 The Bilingual Lexicon	56
6.2 Transfer Effects as Overgeneralisations	58
6.2.1 Paradigmatic Identity	59
<b>7. Concluding Remarks</b>	<b>60</b>

## **CHAPTER 2: Transfer and Delay: Two Sides of the Same Coin?**

<b>1. Introduction</b>	<b>67</b>
<b>2. Transfer in WH-Elements</b>	<b>68</b>
2.1 An English-Italian Case Study	68
2.1.1 Introduction	68
2.1.2 The Subject	69
2.1.3 The Data	70
2.1.4 The Updating Process	72
2.1.5 Some Remarks on Monolingual Transfer	73
2.1.6 Summary	77
2.2 Cantonese v. English	77
2.2.1 Typology	77
2.2.2 Paradigm Formation	78
2.3 A Cantonese-English Case Study	79
2.3.1 Introduction	79
2.3.2 The Data	79
2.3.3 Remarks on Unidirectionality	81
2.4 A Second Cantonese-English Case Study	82
2.4.1 Introduction	82
2.4.2 The Data	83
2.5 Monolingual Cantonese	84
2.5.1 Introduction	84
2.5.2 The Experiment	85
2.5.3 Results	86
2.5.4 Analysis	86
2.6 Summary	88
<b>3. Negation and Modality</b>	<b>89</b>
3.1 Transfer of Modality	89
3.1.1 Introduction	89
3.1.2 Background	89

3.1.3 Data Analysis	91
3.1.4 A Case for Transfer	94
3.1.5 Summary	99
<b>4. Transfer as Delay</b>	<b>99</b>
4.1 Introduction	99
4.1.1 The Data: Spontaneous Speech	100
4.1.2 The Data: Repetition Task	102
4.1.3 Theoretical Considerations	104
4.2 Negative Concord	104
4.2.1 N-Words as Negative	106
4.2.2 N-Words as Existentials	107
4.2.3 N-Words as Indefinites	108
4.2.4 N-Words as Universals	112
4.3 N-Words: The Children's View	116
4.3.1 Introduction	116
4.3.2 Italian Input	118
4.3.3 English Input	120
4.3.4 Lexical Representations	122
4.4 Summary	123
<b>5. Concluding Remarks</b>	<b>124</b>
 <b>CHAPTER 3: The Importance of Being a Paradigm</b>	
<b>1 Introduction</b>	<b>132</b>
1.1 Some Current Issues	133
<b>2. Delay of Principle B Effect</b>	<b>134</b>
2.1 The Phenomenon	134
2.2 Absence of DPBE: Greek	137
2.3 Greek Pronouns	138
2.4 Implications for Transfer	141
2.4.1 Feature Specifications	141
2.4.2 Addressing Transfer	144
<b>3. DPBE and Bilingualism</b>	<b>145</b>
3.1 Greek-Dutch Bilinguals	145
3.2 Greek-English Bilinguals	146
3.3 Interlude	147
<b>4. DPBE in Monolinguals</b>	<b>148</b>
4.1 Absence of DPBE: Clitics	148
4.1.1 Introduction	148

4.1.2 Experimental Evidence	148
4.1.3 Syntactic v. Phonological Clitics	149
4.1.4 Monolingual Transfer	151
4.1.5 Clitic Placement	151
4.1.6 Serbo-Croatian Pronouns	153
4.2 Beyond Generalisation	154
<b>5. Implications for Overgeneralization</b>	<b>158</b>
6.1 The Role of Lexical Specification	158
6.2 The Experiment	159
6.2.1 The Subjects	160
6.2.2 Material and Procedure	160
6.2.3 Control Conditions	161
6.2.4 True-Grammatical	162
6.2.5 False-Ungrammatical	162
6.2.6 Results	162
<b>6. Conclusions</b>	<b>164</b>
 <b>CHAPTER 4: On the Immunity of Pro-Drop and Root Infinitives</b>	
<b>1. Introduction</b>	<b>167</b>
1.1 More on the Lack of Transfer	168
<b>2. The Case of Pro-Drop</b>	<b>169</b>
2.1 Remarks on Richness	170
2.1.1 Introduction	170
2.1.2 Feature Binarity	171
2.1.3 Richness as Opposition	172
2.1.4 How Poor is Rich?	174
2.1.5 The Irish Paradigm	177
2.1.6 Explaining Koenenman and Neeleman's Generalisation	179
2.1.7 Speculations on Opposition	181
2.1.8 Interim	183
2.2 Consequence for Cascade-Updating	184
2.2.1 Transfer and Pro-Drop	184
2.2.2 Morphological Transfer	184
2.2.3 Interim	187
2.3 Evidence from Bilingual Studies	187
2.3.1 Catalan-English	187
2.3.2 Italian-English	188

2.3.3 Inuktitut-English	191
2.3.4 Interim	193
<b>3. Root Infinitives</b>	<b>193</b>
3.1 The Data	193
3.2 Infinitives, Agreement and Pro-drop	196
3.2.1 Consequences for Cascade-updating	197
3.3 Evidence from Bilingual Studies	198
3.3.1 German-Italian	198
3.3.2 German-Italian 2	199
3.3.3 An Alternative Analysis	200
3.3.4 English-Spanish	203
3.3.5 What about AgrS?	205
<b>4. Concluding Remarks</b>	<b>206</b>
 <b>CONCLUSIONS</b>	 <b>212</b>
 <b>References</b>	 <b>217</b>
 <b>Appendix</b>	 <b>236</b>

## Acknowledgements

Bringing a doctoral degree to completion is very much like a pilgrimage, and pilgrims collect a number of debts on their way. Mine are mostly intellectual debts, though not entirely so.

In the first place, I am grateful to my supervisor Hans van de Koot for seeing me through this journey and sticking with me in all weathers. Even when the road got rough, his patience and commitment made sure that I kept walking forward.

My most heartfelt thanks to my second supervisor Neil Smith whose immense knowledge and profound understanding have made the good patches great and the bad ones palatable. The fact that it was Neil who first introduced me to the subject of language acquisition is also of some importance, though he should not be held responsible for what I have done with it.

The research presented in this thesis could not have been carried out without the help of the staff and children of the Scuola Materna A. Palli in Voghera (Italy). I wish to thank all them for their hospitality and kindness. Thanks are also due to Adriana and Massimo Gamalerio for welcoming me into their home, where their son Lorenzo produced more interesting utterances than I could possibly hope to discuss. Special thanks also to Alex Perovic for sharing with me her experience on the experimental procedure and material. I also acknowledge financial help from the Arts and Humanities Research Council (award no: 03/104569) and the Old Students Association (award no: NG4654 – 116/776) for partially funding this research.

A number of linguists have been kind enough to discuss some of the ideas and worries that accompanied this research, these are Dirk Bury, Olaf Koeneman, Ad Neeleman, Kriszta Szendrői, Hiroyuki Uchida, Spyridoula Varlokosta, Deirdre Wilson, Anita Wong, Yi Xu. Thanks to all of you. Special thanks go to Andrew Radford and Vieri Samek-Lodovici whose insightful comments have contributed to the final version of this thesis.

A big thank you also to Nick Allott and Vikki Janke, two linguists who kept away from my theoretical worries but who, on more than one occasion, have offered me a hand out of intellectual swampland. My thanks also to Sam Martin for being an inspiring mentor and for constantly having reckoned with the job of reminding me that research is but a satellite in the universe of life.

Finally, I am forever indebted to Eirini Sanoudaki for the time and patience with which she has sat listening to my endless conjecturing and for being an extraordinary human being.



# INTRODUCTION

---

This thesis investigates some of the fundamental issues within the field of bilingual first language acquisition, the area of research that studies the linguistic development of children who are exposed to two languages from birth.

Various studies on bilingual first language acquisition (henceforth BFLA) have investigated the possible ways in which the language system deals with being exposed to input from two different languages.

A particular phenomenon that has been central to BFLA, and the one which will be the focus of this thesis, is that of “transfer” or “interference” effects. These are cases where bilingual children use apparent literal translations, where lexical items from one language are structured as they would be in the other. The following is an example from a German-Italian bilingual (from Volterra and Taeschner 1978):

1. Lisa will nur Schuhe dunkelbraun (Lisa 3;6)  
Lisa wants only shoes brown

The sentence in (1), which is ungrammatical in adult German, features the exact syntactic structure of Italian, Lisa’s other L1.

For decades, this phenomenon was the subject of much debate between two major theories that arose from the psycholinguistic tradition: the Separate Development Hypothesis (Genesee 1989, Meisel 1989, among others) and the Single System Hypothesis (originally developed by Volterra and Taeschner 1978). While the former claimed that the two languages develop completely separately from day one, the latter defended the idea that the child develops a so-called “hybrid system” which is a combination of the two target languages. Perhaps surprisingly, most of the work done within these two traditions showed virtually no interest in issues concerning the architecture of language or the acquisition process. Researchers focused almost exclusively on whether the child does or does not

separate the two languages during acquisition without ever addressing the question of why either case should apply at all.

Moreover, this debate kept the focus away from a characteristic of BFLA which I believe to be of major importance, namely the fragmentary nature of transfer effects. As soon as this is considered it becomes evident that, strictly speaking, neither of the two traditional hypotheses can be correct. If it were the case that the two languages are completely separate, the child's utterances should feature absolute absence of any apparent translations unless these are a consequence of a particular developmental stage within one language, in which case they will also be present in the speech of monolinguals. On the other hand, if the two languages gave rise to a hybrid system, transfer effects should be present indiscriminately across the board and in copious amounts. Neither of these scenarios reflects the facts. For example, De Houwer (1990) showed that the subject of her case study, Kate, used each language in a manner largely consistent with the language of her interlocutors, though she arguably displayed some transfer effects. In the same vein, Genesee (1989) discussed a series of experiments which showed, rather conclusively, that although bilinguals display occasional transfer effects, they systematically use each language predominantly with speakers of that language (1989). Furthermore, Genesee et. al. (1995) reviewed a number of cases showing that the path bilingual children follow during the development of their two languages closely resembles the one followed by their monolingual peers.

This type of evidence has been largely taken to indicate that the Single System Hypothesis is untenable, although I believe it also raises questions about the defensibility of the Separate Development Hypothesis. Most of all, there seems to be a necessity for formulating the idea of separation in a manner which allows some type of contact between the two languages. However, a precise reformulation of the Separate Development Hypothesis is yet to be provided.

Nevertheless, as the Separate Development Hypothesis gained wider support, researchers have taken a different perspective on BFLA. For example, Döpke (2000) points out that the sporadic nature of transfer effects should not serve as an excuse for disregarding them, since they could be invaluable clues as to the cognitive processes involved in the simultaneous acquisition of two languages. In the same vein, Müller (1998) suggests that, although the two languages may develop separately, they still may have some influence on each other and that it is the domain of the bilingual researcher to investigate how this influence process might function. A new and more specific question is thus raised with regard to the

architecture of the acquisition process in general and how it can result in the development of two separate, yet connected, language systems.

The aim of this thesis is to contribute towards an understanding of these issues by proposing a potential explanation for the occurrence of transfer effects and investigating the theoretical and empirical consequences of this explanation.

Firstly, I will be concerned with the structure of the acquisition device. The bulk of chapter 1 will be dedicated to this purpose. Based on assumptions that are mostly drawn from standard linguistic theory, I will propose a theoretical model of language acquisition which might shed some light on the workings of bilingual acquisition and enable us to make some predictions as to what particular structures may be subject to transfer effects. In particular, I will argue that the mechanism responsible for the emergence of transfer effects is the very same one that underlies the well-documented overgeneralisation process in monolinguals. This link between transfer effects and overgeneralisation is arguably the main motivation behind the model to be discussed in this thesis.

The necessity to link BFLA to monolingual acquisition has sometimes been acknowledged elsewhere in the literature. For example Hulk and Müller (2000) suggest that transfer is “more likely to occur in exactly those areas which are also problematic - albeit to a lesser extent - for monolingual children” (2000:228). This statement refers to their idea that bilingual and monolingual children alike have ‘problems’ with structures usually associated with the C-domain (e.g. verb second, complementizers, and topicalization). According to this view, transfer effects result from a ‘relief strategy’ which bilinguals are said to apply in order to deal with the problems that arise from the alleged vulnerability of the C-domain.

Nevertheless, I believe that if a theory of BFLA is to achieve explanatory adequacy it must necessarily regard transfer effects as a consequence of the workings of the language acquisition process. Hypotheses based on claims that bilinguals employ some ‘special’ principles or ‘relief strategies’ are undesirable as they only provide a formal description of the problem. Moreover, if the language behaviour of bilinguals results from the application of special principles then bilingualism is of little interest to the acquisitional theorist since it sheds no light on the workings of the acquisition process per se. On the other hand, if we take such language behaviour to be a direct consequence of the very system that underlies acquisition, its investigation will contribute to our understanding of that system. Taking this as a point of departure, I will try to derive transfer effects from the mechanisms that underlie language acquisition in general. This will lead us to the second aim of this thesis, namely that of assessing the empirical consequences of

the acquisitional model to be proposed. This will involve discussion of various data-sets from the literature as well as of newly conducted research and will be the focus of the remaining three chapters, to be organised as follows.

In chapter 2 I will introduce examples of transfer effects from two case-studies which involve two different language pairs, namely English-Italian and Cantonese-English. I will then show how these can be explained by appealing to the architecture presented in chapter 1. The same chapter also involves discussion of some non-adult utterances which - *prima facie* - appear to be manifestations of two different phenomena, namely transfer effects and delay. I will suggest that - contrary to traditional views - these do not involve separate cognitive mechanisms and can be accommodated as cases of transfer, despite the fact that they appear to be radically different on the surface. Some issues concerning overgeneralisation in monolinguals will also be addressed in this chapter.

In the third chapter I will discuss some experimental evidence which indicates that transfer effects do not affect the development of Principle B of Binding Theory (Chomsky, 1981). I will then argue that this follows from the architecture proposed in chapter 1. I will also present some newly obtained experimental evidence in support of the same analysis but with regard to monolingual development.

The fourth chapter discusses two other areas within which transfer effects have been argued to be absent, namely Pro-drop and Root Infinitives. I will argue that this is in fact predicted by our analysis since these phenomena are related to principles that reside outside the domain of application assumed for our model.

Chapter 5 gives some general conclusions. An appendix is also included with details of the experimental material employed as part of the experiment discussed at the end of chapter 3.

# CHAPTER 1

---

## On the Acquisition of Lexical Properties

### 1. Introduction

Developing a theory of language acquisition is a massive undertaking and I am not hoping to address every aspect of language development, or of the mechanisms involved in it. However, it is my intention to present a view sufficiently precise to help us make some headway in the domain of lexical generalisation and, ultimately, transfer effects.

The aim of this chapter is threefold. Firstly, I will introduce a particular architecture of the language system, based on the work of Jackendoff (1997). I will then outline a specific model of language acquisition which involves updating of lexical items followed by systematic generalisation of newly acquired properties. Finally, I will show how this may help us understand some well-known cases of overgeneralisation and suggest that the same system can also provide an explanation for the transfer effects found in bilinguals. More precisely, I will argue that given certain assumptions about the organisation of the lexicon, transfer effects are a necessary consequence of the acquisition process.

### 2. The Acquisition Process

#### 2.1 Preamble

In addressing the issue of language acquisition I will assume that humans are endowed with innate linguistic knowledge in the form of a Universal Grammar (Chomsky 1962, 1981). This assumption will provide an essential scaffold on which a potentially successful theory of language acquisition can be developed. In particular, it will play a fundamental part within the context of UG-compatibility, a

notion which will become central to our discussion when addressing well-known learnability issues (cf. section 2.4 and section 5 below).

Following a theory of UG necessarily implies assuming that there are some linguistic components which are built into what Chomsky (1986b) calls the *initial state* of language acquisition. However, there has been much debate in the literature with regard to how much should be assumed to be available to the child. A variety of positions have been advocated, often within either a Continuity (see for example Pinker 1984, Poeppel and Wexler 1993, Wexler 1999) or a Maturational perspective (Borer and Wexler 1987, Radford 1988, 1990, Rizzi 1994, Wexler 1994, among many others). Both Continuity and Maturational views come in a variety of formulations. In its strongest formulation, the Maturation hypothesis claims that language development involves not only the acquisition of lexical items but also the addition or development of some specific categories and principles. For example, Radford (1990) argued that the relative absence of elements such as determiners and auxiliaries in early child speech can be explained by assuming that functional categories are not represented in early child grammar and that these mature at a later stage (see also Meisel 1994, Platzack 1990, Tsimpli 1996). A different maturational account has been proposed by Rizzi (1993/1994) and Wexler (1994, 1996) who suggested that children's linguistic behaviour does not necessarily lead to the conclusion that functional categories are absent at the early stages. Instead, they suggest that what needs to be assumed is the maturation of some principle that disallows optionality. In Wexler's terms, this would be a principle which dictates that Tense must be obligatorily projected. Similarly, Rizzi suggests maturation of a principle stating that root clauses must necessarily consist of a full Complementiser Phrase. What sets these proposals apart from Radford's is their commitment to the claim that the initial state includes all the syntactic categories allowed by UG, including functional categories.

As for the Continuity hypothesis, there are at least two formulations of it, often referred to as 'weak' and 'strong' continuity, respectively. Strong continuity accounts maintain that all UG properties and principles are available to the child from the initial state. The obvious differences that exist between child language and the adult target are put down to putatively UG-external factors. These include constraints on phonological production (Demuth 1994, Gerken 1994), interface problems between different linguistic components (e.g. Phillips 1996), or processing limitations (Poeppel and Wexler 1993). However, it is not obvious why some of these alleged solutions should qualify as UG-external. For example, it is at best

unclear why some phonological constraints may be assumed to be subject to developmental stages while syntactic ones may not.

It is perhaps with this kind of objection in mind that some researchers have opted for a weaker formulation of the Continuity hypothesis. This involves assuming that although properties and principles do not mature, they may be present in an underspecified version when compared to the target language. In other words, the fact that child language may deviate from the target is explained solely in terms of parameter/feature setting. This position is taken by, among others, Déprez (1994) who suggests that optionality in child grammar can be explained by assuming that functional categories may be underspecified in early child language (see also Radford 1995, Vainikka 1993/1994). Since under-specification is a possible setting in adult grammars too (see also section 2.1 below), it can be maintained that child language is consistent with adult language, though not necessarily with the target. In other words, every instance of child language corresponds to a possible adult language. Consequently, the weak continuity hypothesis is largely compatible with most versions of Maturation, as most maturational accounts assume that there is no principle which is peculiar to child language (see for example the maturational views proposed in Borer and Wexler (1987) and Radford (1988, 1990)).

At first glance, it seems inevitable that any theory of acquisition must necessarily be developed with one of these assumptions in mind. Nevertheless, in this thesis I will abstract away from the Continuity/Maturation debate as I believe that the model I will be defending is compatible with either assumption, for the following reasons. The focus of this thesis is on how the language acquisition device deals with a lexical feature/property once this has been associated with some items (cf. section 2.4.3) and the impact this has on the rest of the lexicon (cf. section 3). Whether the acquisition of the feature/property at issue is dependent on maturational principles or on triggering experiences does not seem to be of consequence. Because the model I will be presenting only makes claims about the mechanisms that affect a feature/property once this comes to be hosted by some lexical items, what happens before this process takes place (i.e. the feature is immature, dormant, below threshold etc.) does not affect our line of argumentation. Indeed, some of the assumptions I will be making (e.g. UG-compatibility, cf. section 2.4.5) have been previously made by proponents of both Continuity and Maturation.

At the same time, as the discussion develops I will suggest that certain specific properties should be assumed to be present from the initial state (see for example sections 2.2 and 3.2.1 with regard to person features and conceptual structures respectively). This is not equivalent to supporting the Continuity

hypothesis, as Maturation approaches also postulate the existence of certain properties from the initial state. Nevertheless, it does entail that the model I propose may be incompatible with certain specific Maturation accounts. For example, the discussion in section 2.2 suggests that the model to be proposed is incompatible with any account that assumes maturation of person and number features (though I am not aware of any work within the universalist tradition that is in line with this assumption, thus the above point may be a trivial one).

In sum, in those cases where I will be arguing that certain properties are present from the initial state, I will be defending the position that there is developmental 'continuity' as far as those properties are concerned. This view is compatible with the Continuity hypothesis as well as with a number of Maturation accounts, since lack of maturation for a given group of properties does not contradict Maturation *per se*, it simply suggests that such properties do not mature, while others still may.

A more immediate question, and one which is not necessarily dependent on the Continuity/Maturation debate, regards the identity of the properties that might be present from the initial state of language acquisition. I will address this in section 3.2. First, however, I will lay out some assumptions concerning the architecture of the acquisition process as well as the nature of the mechanisms involved.

### 2.2 Lexical Properties and their Value

During acquisition, the crucial task for the language system is to detect the presence of a particular feature/property in the input and, consequently, to work towards its acquisition. Following much current work I will assume that lexical features are of a binary nature. With regard to this, some clarification is in order. It has sometimes been suggested (e.g. Adger 2003, Brody 2002) that the most economical distinction we can postulate is of a monadic nature, i.e. based on the *presence* vs. *absence* of a feature. Although a possible representational choice, this assumption raises a serious learnability problem. Consider a case in which the acquisition device has concluded that some lexical item has a negative setting for a feature *f*. Within a monadic system, this is equivalent to acquiring absence of *f*. This process offers no obvious way of preventing the acquisition device from considering addition of *f* again and potentially re-acquiring the same *f* feature which it had previously concluded should be absent. Therefore, it seems that a monadic system is unable to recognise whether it has achieved a target setting, namely *absence* of *f*, or whether it has not yet been exposed to such feature at all. The only way of enabling such a distinction is to introduce some memory device whereby deletion of



*f* is perceived differently from absence of *f*, essentially introducing a binary system. I will briefly return to this in section 5 (see also Fodor and Sakas 2005 for extensive discussion of how memory-less acquisition systems are bound to fail). I therefore embrace the view that a feature must be set to either +*f* or -*f* and I will take the absence of a feature as indicating the system's lack of knowledge with regard to that particular feature.

As has been convincingly argued for within morphology (Ackema 2001, Andrews 1990, Blevins 1995), however, this cannot be the full picture since the presence of a feature does not necessarily entail the presence of a specific value. In particular, Blevins (1995) – building on the work of Andrews (1990) – proposes that morphological syncretism should be taken as arising from feature underspecification. The features associated with each paradigm member, he argues, must necessarily be defined in direct opposition with other members if feature-based analyses are to move beyond simple descriptive formalism. The English verbal paradigm illustrates this point:

1. English Inflectional Paradigm

walk	1-sg	walk	1-pl
walk	2-sg	walk	2-pl
walk- <b>s</b>	3-sg	walk	3-pl

Besides creating some unnecessary redundancy, associating the form *walk* with five separate feature specifications is conceptually unrevealing to say the least. A much more parsimonious alternative is to recognise that the two separate forms stand in direct opposition:

2. Paradigmatic Opposition: English

walk- <b>s</b>	marked: 3-sg
walk	unmarked, general form

It can then be assumed, following Andrews (1990), that a condition on morphological blocking will prevent the general form from applying when a more specified one exists<sup>1</sup>, hence the ungrammaticality of examples like *\*he walk*.

A natural formalisation of this idea would be to assume that the language learner applies some form of the Biuniqueness Principle (Dressler 1985) which entails that every morpheme corresponds to only one feature specification and every feature specification to only one morpheme. Within morphology, the mapping

of a single phonological string onto two separate feature bundles (cf. 1) is disallowed, as is the co-existence of two separate phonological strings mapped onto the same feature bundle<sup>2</sup> (hence preventing the formation of *he walk* where *walk* would have the same specification as *walks*<sup>3</sup>). These are illustrated by (3a) and (3b) respectively:



The emergence of general forms can therefore be viewed as a consequence of the principle in (3a). The principle in (3b), on the other hand, is a formalisation of the blocking principle mentioned above and it is responsible for the absence of free variation between specified (e.g. *walks*) and general forms (e.g. *walk*). Several formulations of this principle have been proposed in the morphological literature (see, for example, Aronoff 1976, Lapointe 1980, Pinker 1984)<sup>4</sup>.

Another important reason against the analysis in (1), and therefore in favour of the principle in (3a), comes from learnability considerations. Pinker (1984) argues that unconstrained proliferation of feature specifications to be mapped onto homophonous strings leads to serious learnability problems<sup>5</sup>. A theory that allows zero-inflection to occur indiscriminately has no way of preventing children from hypothesising grammaticalised features for every single lexical item that belongs to a potentially inflectable category. Needless to say, this leads to massive computational waste and potential failure to achieve the target language.

Consequently, Pinker suggests a development of Slobin's (1984) hypothesis and proposes that postulation of feature specifications must be limited to those cases where the input provides a morphological contrast in the form of some overtly realised feature bundle(s), as is the case in the English verbal paradigm. This would successfully solve the computational problem. Because the child encounters the *-s* affix in third person singular cases, s/he will be able to postulate a specific 3<sup>rd</sup> person singular form and contrast this with an underspecified 'elsewhere' case. In cases where no morphological marking is present at all, however, there is no contrast to trigger the postulation of specific features. Therefore, the number of specific features postulated by the language learner is directly proportional to the number of distinct morphological forms found in the input (see also Koenenman 2000). Application of the morphological principles in (3) further simplifies the learner's task of mapping morphological content onto phonological strings.

Inevitably, however, the claim in (2) raises the question of how this kind of generality is formally encoded. One possibility is that of analysing generality as a consequence of feature disjunction (cf. Karttunen 1984). On this view, the English verbal paradigm would look as follows:

4. Generality as disjunction

walk- <b>s</b>	3-sg
walk	1-sg $\vee$ 2-sg $\vee$ pl

This is hardly an improvement over analyses such as (1) above as it allows the amalgamation of features that do not form a natural class, leading to essentially “arbitrary formal specifications” (Blevins 1995:124)<sup>6</sup>. This type of analysis also fails to capture an important crosslinguistic generalisation, namely that syncretism does not occur randomly. For example, Williams (1981) points out that – in the nominal domain – syncretism is widely attested between nominative singular and accusative singular but not between nominative singular and accusative plural (see also Carstairs 1983).

A more promising alternative is to analyse the general case as *underspecified* for Person and Number features, which gives the opposition shown below<sup>7</sup>. Following a long tradition (dating back to Chomsky and Halle 1968), alpha-notation is used to indicate underspecification:

5. Underspecification

walk- <b>s</b>	3-sg
walk	$\alpha$ person, $\beta$ number

This offers a satisfactory solution to overspecification problems that arise with analyses such as the one given in (1) above without having to resort to arbitrary specifications.

I will therefore adopt the view that features can be *underspecified* as a result of the application of the Biuniqueness Principle which I will assume to be at work within morphological paradigms (but see Leiss 1997 for a different position which advocates the Biuniqueness Principle being active across grammatical categories).

We are now ready to lay out what possible feature settings may be relevant to the acquisition process. These correspond to either [+f] or its counterpart [–f] as well as the underspecified [ $\alpha$ f]. Crucially, the underspecified case cannot be taken to result from the total absence of a feature since, as mentioned above, total

absence of a feature indicates that such feature has not yet been acquired. I will therefore make use of the notation  $[af]$  to indicate that a feature is underspecified.

In sum, there are four distinct states of affairs that can obtain with regard to any given feature:

- |    |    |  |
|----|----|--|
| 6. | a. | no knowledge of $f$                                    |
|    | b. | $af$ Knowledge of $f$ but underspecified value for $f$ |
|    | c. | $+f$ Knowledge of $f$ specified as +                   |
|    | d. | $-f$ Knowledge of $f$ specified as -                   |

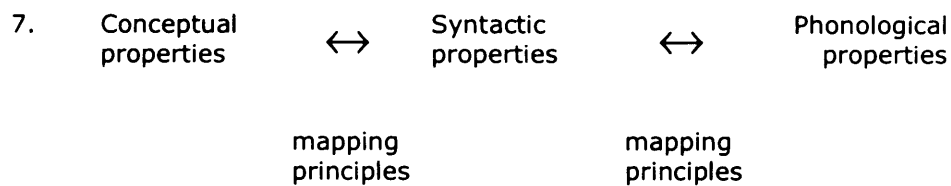
The cases represented in (6) will have the same effect on language production independently of whether they obtain during acquisition or as part of the adult knowledge. However, the states represented in (6a) and (6b) can have different implications due to the different status that the child's and the adult's system have with regard to accessibility. As far as the child is concerned, absence of  $f$  and underspecified  $f$  can be two subsequent stages that precede a complete setting for either  $[+f]$  or  $[-f]$ . If a feature is temporarily absent in the child's knowledge, it might later be detected in the input and therefore be made part of such knowledge. When this happens, however, the child might not yet have derived the value for it. S/he will therefore set it as underspecified (i.e.  $[af]$ ) until further evidence is gathered and a value can be set. Indeed, it could turn out that the target language has no specification for  $f$  and therefore there will be no further value to add. As for the adult, the four cases represented in (6) are independent of each other, having no chronological connection.

### 2.3 Language Variation

Following the generally accepted assumption that language variation is restricted to the lexicon (Borer 1984, Elliott and Wexler 1985, Chomsky 1995) I will view language acquisition as the process involving the setting of the idiosyncratic combination of properties that makes up the identity of each individual lexical item. On this view, acquiring a lexical item is equivalent to setting the semantic, morpho-syntactic and phonological properties that will determine the item's behaviour within each of the linguistic components.

Following Jackendoff (1987, 1997) I will take these components to be Representationally Modular. This view takes the three components as independent of one another, each with its own set of "combining principles" and distinct "representational formats" (Jackendoff 1997:41). Consequently, grammaticality

involves well-formedness of a derivation within the specific component that generated it (syntax for a syntactic structure, phonology for a phonological structure etc.), as well as successful mapping onto the other components through mapping principles applied at the interfaces. Within this model, lexical information takes the form of spell-out rules whose function is to associate conceptual and syntactic representations with phonological material. These have the following representational structure:



While the workings of each component are universal, mapping principles may vary from language to language, thus yielding language-specific idiosyncrasies. The immediate advantage of this view is that the linguistic components are relatively simple and complexity lies at the interface level with the mapping principles.

It is important to note that the architecture in (7) represents a radical departure from traditional views of the lexicon. Firstly, it views the lexicon as a group of associations rather than a repository for lexical items (see also section 3.7). Secondly, these associations do not necessarily correspond to terminal nodes, since phonological information is associated with (syntactic and semantic) representations rather than with elements<sup>8</sup>.

The idea that lexical mappings may also involve non-terminal nodes has been around at least since Wells (1947), though its importance has also been highlighted within generative frameworks. Consider the following examples, from Radford (1988:175):

- 8.
- a. The present [king of England] is more popular than the last *one*
  - b. \*The [king] of England defeated the *one* of Spain

The contrast between (8a) and (8b) can be easily captured by positing that *one* realises N-bar rather than N or NP. Although this may pose a problem for more traditional lexical models, it can be straightforwardly accommodated within a representationally modular framework as this allows phonological words to be mapped onto non-terminal nodes (we will return to this point in chapter 3, section 2.4). The need for models that allow this type of mapping has also been pointed out

in recent work by Weerman and Evers-Vermeul (2002) and Neeleman and Szendrői (2005) based on evidence from Dutch.

Returning to acquisition, a representationally modular system involves two core processes, as outlined in (9).

9. a. Acquiring combinations of properties relevant to each separate component
- b. Mapping those combinations of properties onto representational elements of other components.

In what follows I will particularly focus on how the task in (9a) might proceed, though I will also examine various aspects of (9b) as the discussion develops.

## **2.4. Acquiring Language-Specific Idiosyncrasies**

### **2.4.1 Introduction**

By its very nature, the acquisition process involves constant updating of lexical information based on the incoming input, until the system contains the same combinations as the target language. Let us consider how this might proceed. As the child is exposed to linguistic input, the language system is provided with evidence that a particular property *P* is associated with a lexical mapping *L*. Such evidence will cause the language system to begin working towards updating *L* by adding *P*.

As a working example, let us consider the acquisition of I-to-C movement in English auxiliaries. Following a suggestion by Rizzi (1990, 1996), suppose that this type of movement is triggered by a universal requirement that two elements bearing a [+wh] feature must be in a spec-head relation at some point in the derivation. Suppose further that, in the target language, tensed auxiliaries are associated with a [+wh] feature and must therefore raise to the  $C^0$  position when a wh-element is inserted into spec-CP (i.e. in interrogative contexts)<sup>9</sup>. On this view, acquiring I-to-C movement is thus equivalent to associating the [+wh] feature with those elements that bear the features [+V, -N, +A]. At first, the child will do so on an item-by-item basis. On hearing (multiple tokens of) an affirmative/interrogative pair *NP can INF/can NP INF?*, the child will associate [+wh] with the modal *can*, provided that s/he has already acquired the relevant structures for interpreting the input at issue. Within a Representationally Modular framework, this process results in the updating of a spell-out rule, thus moving from the configuration in (10a) to that in (10b). Boldface indicates a newly acquired feature/property<sup>10</sup>:

10.

- a.            [+V, -N, +A]                            ↔            /kæn/
- b.            [+V, -N, +A], **+wh**                            ↔            /kæn/

Importantly, however, this process cannot be assumed to proceed piecemeal for each modal auxiliary. Although the child will begin by acquiring the [+wh] specification based on exposure, s/he will eventually generalise by associating this type of feature with all auxiliaries. Crucially, this involves associating the [+wh] feature with elements for which the child may not have had evidence of movement.

Examples of this ability to generalise abound. For instance, it has often been reported that production of modified (and definite) NPs in young children (MLU 1;0 – 2;0) is limited to the object position, presumably due to some association of modification with accusative case (Bloom, Lightbown, and Hood 1975, Brown 1973). However, as soon as the first modifiers appear in conjunction with nominative NPs, there is a tremendous surge in the amount of modified nominative NPs being produced. A very similar pattern has been reported for the acquisition of control verbs. Acquisition, therefore, does not involve learning all lexical combinations individually. Once a certain property has developed, it is systematically associated with different structures. Note that this cannot be captured by appealing to universality (i.e. UG). Admittedly, the systematicity of a process may sometimes result from the application of universal principles (an issue to which we will return in section 2.4). Nevertheless, the amount of cross-linguistic variation with regard to – for example – which particular verbs allow for a control structure, indicates that some of these processes will inevitably involve language-specific mappings.

Within the generative tradition, language-specificity has often been dealt with by postulating some form of rule-formation process often assumed to apply within the lexicon (see for example Jackendoff 1975, Pinker 1989). On this view, the language learner's task was partly to build language-specific output rules. However, this is incompatible with a Representationally Modular system<sup>11</sup> that regards language acquisition as the pursuit of the goals in (9). This is because the suggestion that language acquisition also involves the formation of output rules presupposes a system whereby languages can vary as to the rules they apply.

Nevertheless, there is something about the language system that enables it to capitalise on newly acquired knowledge and any potentially successful model of acquisition must somehow reflect this. In the following sections I will propose a

particular formalisation of this effect which is compatible with the model assumed so far. The central claim of this proposal is that there exist two interacting 'higher level' systems that can fully access lexical information within each of the modules, as well as the mapping relations between them (though not necessarily the inner workings of the modules, see section 2.4.4). The first of these systems, which I will call *paradigm formation*, is dedicated to grouping lexical mappings based on the lexical properties they host. The second system, I will argue, is an *informationally monotonic* updating system whose role is to add newly acquired lexical information to those items that are not yet fully developed. I will present them in turn.

### 2.4.2 The Notion of Paradigm

The term "paradigm" is heavily loaded with theory-specific connotations, the strongest of which is perhaps that of morphological relations (as used in section 2.2 above). As far as acquisition is concerned, Pinker (1984) discusses a process whereby children must unravel morphologically complex words by associating each affix with the specific environment in which it occurs and – ultimately – with a specific cell within a matrix representation (i.e. the paradigm). Within the computational tradition, on the other hand, the term "paradigmatic relation" is often used to refer exclusively to semantic relations, specifically meaning relations such as synonymy, polysemy and so forth (see, Kahane and Polguere 2001, Gaume et. al. 2002, among others).

In formulating my view of what constitutes a paradigm I would like to take a step back and reconsider a much older insight, on which – I suspect – the above analyses were also based, namely that of Saussure (1949). In its original formulation, this view of a paradigmatic relation was concerned with elements that belonged together by virtue of a function they shared. This was initially gauged in terms of substitution. Translated into somewhat more modern terms, this view takes two elements as paradigmatically related if they can be freely substituted for each other without affecting the grammaticality of a string. What I would like to suggest is that the study of language acquisition has much to gain from pursuing this idea within a formal grammatical system and I hope I can show how this may be done.

Roughly speaking, I would like to suggest that two elements are in a paradigmatic relation if they share at least one formal property. This differs typologically, if not in spirit, from the more common uses mentioned above. Specifically, it does not limit paradigm relations to closed-class elements (*contra*

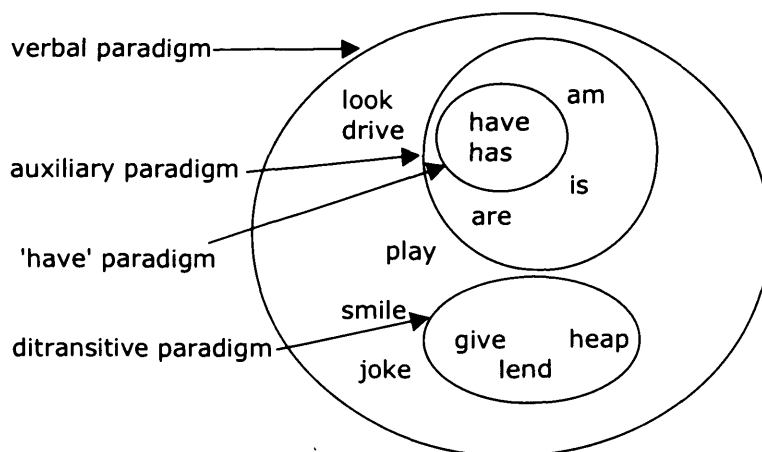


Pinker 1984:166) nor does it suggest that they refer only to semantic information (*contra* computational literature).

Let us take, as a starting point, the notion of word category. Given our tentative definition of paradigmatic relation, all words that share the same word category belong to a paradigm in that they share a particular formal property in the form of their syntactic category. We may therefore talk about a verbal or a nominal paradigm when referring to verbs or nouns. Indeed, there are paradigms within a paradigm. An auxiliary, for example, shares some of its properties with lexical verbs but it also hosts a more specific property – traditionally a [+A] feature – which it only shares with other members of this closed class.

When zooming in on the verbal paradigm, we find that each auxiliary has its own paradigm whereby particular forms relate to particular Number and Person features. In the same vein, lexical verbs also form sub-paradigms. For example, ditransitives are paradigmatically related to each other in virtue of the formal property (or properties) which represent their specific subcategorisation requirements and which they only share among themselves. An informal representation of what this might look like is given in (11).

11.



The largest set encompasses all items that share the formal features [+V, -N] while its subsets are based on other properties that items may share *in addition* to categorial features.

As mentioned above, the fact that auxiliaries host an extra property – namely [+A] – will make them part of a more specific [+V, -N, +A] paradigm. In the same vein, the suppletive morphemes within the same auxiliary display one further

level of specificity since they share not only syntactic properties but also semantic ones, thus having the same 'meaning'. In this sense, a paradigm is a set of items that share certain properties, whatever the identity of those properties might be. The larger the number of properties items share, the more specific the paradigm. This follows from the fact that – all else being equal – the size of a paradigm is inversely proportional to the size of the feature bundle which defines it: the higher the number of formal features required for membership, the smaller the resulting paradigm<sup>12</sup>.

Two clarifications are in order. Firstly, the set relations represented in (11) illustrate a viewpoint which has been at the base of our discussion of paradigmatic relations but which has so far been only implicit, namely that paradigmatic relations are first and foremost categorial relations. Hence, the largest set in (11) is a set encompassing a whole word-category. As all other paradigms are contained within the categorial set, our definition of paradigmatic relations will have to include this hierarchical relation somehow. I will return to this in the next section.

The second point that arises is with regard to the type of set relations contained within the categorial set. Although the diagram in (11) only involves subset/superset relations, there is nothing in the discussion so far which implies that other set relations will not be relevant. Indeed, I do not think that other relations should be excluded. For example, if we consider the verb *have* in its transitive use (as in *John has a cat*), it will be the case that the auxiliary set intersects with the set of transitive verbs. However, the consequences of intersection relations will be left pretty much unexplored, as our discussion will be focusing on the role of paradigm formation in extracting lexical generalisations based on newly acquired properties. Within this procedure, the concept of intersection will actually turn out to be of major importance, albeit at a different level of representation. More precisely, we will be concerned with the intersection of the properties borne by lexical elements within a set rather than with the intersection of sets of items themselves (see section 2.4.3).

In the next section I will provide a formal definition of paradigm formation in set theoretic terms. First, however, I would like to discuss its role in the acquisition process.

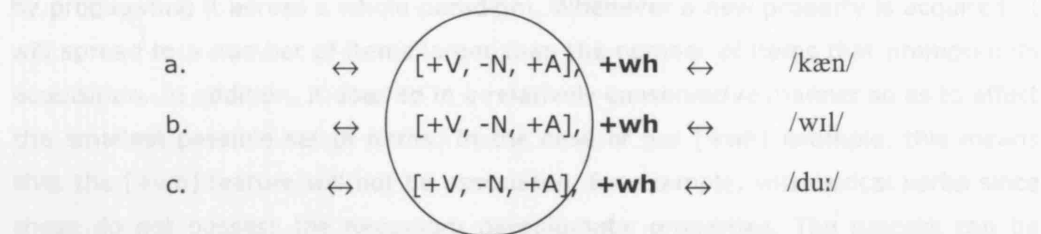
### 2.4.3 Towards a Paradigmatic Acquisition

In section 2.4.1 we saw that when the acquisition system acquires new information on the basis of a number of input items, it applies this knowledge to other items of the same type, independently of whether these contributed to the acquisition of the

new information. Having discussed the nature of paradigmatic relations, we can now begin to sketch out the mechanics of this process. Consider a case where the child has performed the lexical update in (10) for a small set of items, say *can*, *will* and *do*. I will informally call these the trigger elements. Given the lexical map in (11), the language system will be able to identify what specific lexical group is being targeted by the [+wh] feature.

More precisely, given our notion of paradigmatic relations and a set of trigger elements *S*, the language system can deduce a set *P* which contains all those properties that all the trigger elements had in common prior to the arrival of the new feature. This procedure is illustrated in (12), where the circle represents the newly identified set *P* (a set of lexical properties)<sup>13</sup>:

12.



As the only features that all the trigger elements share are the features [+V, -N, +A], these will be recognised as the paradigmatic properties that are being targeted by the newly arrived [+wh] feature. Because we conceive of lexical mappings as sets of properties, the set *P* containing [+V, -N, +A] corresponds to the *intersection* of the sets associated with the trigger elements. Therefore, the procedure that we have so far presented only informally can now be defined via the notion of intersection, as follows:

### 13. Paradigm Formation

Given a list of sets  $S_1, \dots, S_n$ , identify as the paradigmatic properties the set *P* which corresponds to the intersection of  $S_1, \dots, S_n$  minimally containing a full<sup>14</sup> bundle of categorial features.

The setting of a minimal condition is related to the idea that a paradigmatic relation is first and foremost a relation between same-category elements. However, it also serves to prevent vacuous triggering of paradigms based on the empty set.

At this point, the newly acquired [+wh] feature can be automatically associated to all the other sets (i.e. lexical elements) that also contain (i.e. are a

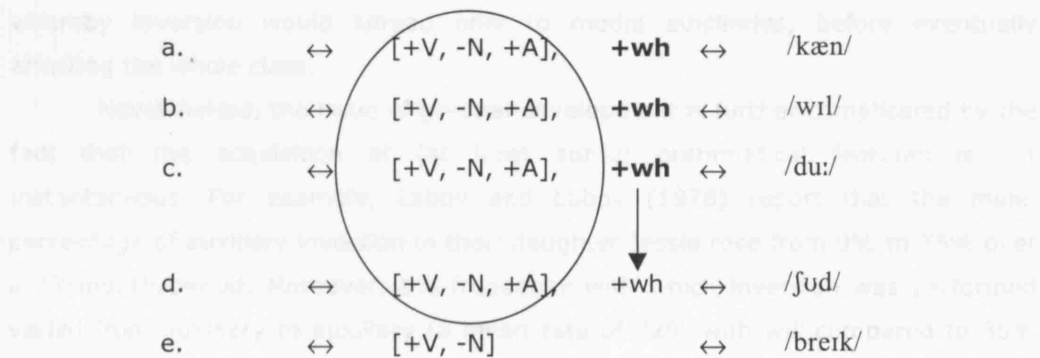
superset of)  $P$ . I propose that this association is the result of the interaction between Paradigm Formation and a dedicated updating system. First, Paradigm Formation detects the paradigmatic properties shared by all the items that host the newly detected feature and then this new feature is automatically assigned to other items that possess such properties. This gives rise to the following cascade effect<sup>15</sup> (initial version):

#### 14. Cascade Updating

A property  $p$  that is newly acquired through the input items  $X_1, \dots, X_n$  will be attributed to item  $Y$  iff  $Y$  possesses the paradigmatic properties common to  $X_1, \dots, X_n$  at the time when  $p$  is acquired.

The advantage of this procedure is that it capitalises on newly acquired information by propagating it across a whole paradigm. Whenever a new property is acquired, it will spread to a number of items larger than the number of items that prompted its acquisition. In addition, it does so in a relatively conservative manner so as to affect the smallest possible set of items. In the case of our [+wh] example, this means that the [+wh] feature will not be associated, for example, with lexical verbs since these do not possess the necessary paradigmatic properties. The process can be represented as in (15), whereby the feature [+wh] is added to an auxiliary verb, but fails to spread to the lexical verb *break* (the arrow indicates acquisition via Cascade-Updating).

#### 15.



As a result of the process in (15) the system will acquire the properties necessary in order to perform inversion of *should* without having to be exposed to any instances

of it. At the same time, the system will not affect lexical verbs, for these do not bear the paradigmatic properties shared by all the trigger elements.

It must also be noted that the architecture I am proposing allows for feature-spreading to occur gradually across a paradigm, and it is thus compatible with cases of gradual development often reported in the literature (Eisenbeiss 2000, Labov and Labov 1978, among others). Consider, for example, a case where the first auxiliaries for which the child has evidence of inversion are *can*, *will* and *could*. If the child has already acquired (some) of the modal properties associated with these elements, the intersection set that Paradigm Formation will extract will necessarily include a modal component. As a result, the feature responsible for auxiliary inversion will initially spread only to modal auxiliaries, as these will be the only elements bearing the relevant paradigmatic properties. Only when the child has evidence of inversion for some non-modal auxiliary will the generalisation spread to the whole +A class. In other words, the Cascade-updating model predicts that a procedure such as auxiliary inversion may develop gradually<sup>16</sup>. At the same time, however, gradual spreading is not a necessary outcome. In some cases, the trigger elements might belong to different subgroups of the auxiliary paradigm, as would be the case if the first items the child learns to invert are *have*, *can*, and *do*. In this case, the intersection set identified by Paradigm Formation would not include any modal properties. The subsequent generalisation would thus affect a wider group of auxiliaries.

Consequently, due to the importance which the Cascade-updating model places on the identity of the trigger items, it is expected that different children may arrive at different generalisations, some of which will include gradual development whereby inversion would spread only to modal auxiliaries, before eventually affecting the whole class.

Nevertheless, the issue of gradual development is further complicated by the fact that the acquisition of (at least some) grammatical features is not instantaneous. For example, Labov and Labov (1978) report that the mean percentage of auxiliary inversion in their daughter Jessie rose from 9% to 76% over a 17-month period. Moreover, the frequency with which inversion was performed varied from auxiliary to auxiliary (a mean rate of 72% with *will* compared to 35% with *can*). Similar findings have been reported with regard to the development of obligatory determiners in German (Eisenbeiss 2000) and Italian (Antelmi 1997), as well as for the disappearance of subject-drop in the speech of children who are exposed to non pro-drop languages (Bloom 1993).

Although I do believe that a model of language acquisition will ultimately need to address the pervasive existence of gradual development, in this thesis I will idealise away from the issue, in the hope that I will be able to address it in future research. For the moment, I will be concerned with the core mechanisms underlying the updating procedure and how these may help us understand some of the cognitive processes involved in acquisition<sup>17</sup>.

#### 2.4.4 *The Role of Paradigm Formation*

The importance of imposing limitations on generalisations can hardly be overestimated (see Baker 1979, Dresher and Kaye 1990, Pinker 1984, among many others). Any learning algorithm that allows generalisations that are too broad will fail on conceptual as well as empirical grounds. Firstly, a system that constantly overshoots the target is conceptually inadequate insofar as it entails massive computational efforts in the form of a continuous “overshoot and retreat” process, subsequently raising questions with regard to its ability to achieve the target. Nevertheless, overgeneralisation is well attested in child language<sup>18</sup> (within some domains, cf. also section 4). This is why any proposal of language acquisition also faces an empirical challenge. On the one hand, a successful model must reflect children’s ability to generalise and simultaneously disallow massive overshooting of the target. On the other hand, it must also allow some overgeneralisation to occur<sup>19</sup> though not beyond what is empirically attested.

Much of the literature within the Principles and Parameters tradition turned to the Subset Principle as a partial solution to this dilemma. This has been formulated in various ways, though the best-known version can probably be found in Wexler and Manzini (1987:44):

16. “Let  $i$  and  $j$  be values of a linguistic parameter  $p$ .  $L(p(i))$  is the language – we take a language to be a set of sentences – which is attained by letting  $p$  have the value  $i$ . Likewise for  $j$ . Then we can state the Subset Principle. Suppose  $L(p(i)) \subseteq L(p(j))$ . Then  $i$  is less marked than  $j$ . In acquisition terms, if  $i$  is less marked than  $j$  then  $i$  is tried before  $j$ , and only positive evidence that  $i$  is wrong moves the learner to  $j$ .”

As extensively argued by Fodor and Sakas (2005), however, the Subset Principle in its original formulation does not provide a plausible learning algorithm. In fact, it is highly unlikely that the Subset Principle is an actual principle, at least in the sense of it being a constituent part of the acquisition device. As transpires from the

definition in (16), the Subset Principle is meant to apply to languages rather than grammars. Consequently, it is only applicable within a system that either (i) has the ability to perform extensional comparisons (and thus to keep track of all the input it meets) or (ii) is programmed with knowledge of all language pairs that stand in a subset-superset relation. It seems undesirable to ascribe either of these conditions to the acquisition device. On the one hand, condition (i) is entirely against the spirit of any generative research program, for it forces onto the learner an ability to memorise sentences (on this point, see Hale and Reiss 2003). Moreover, it begs the question of why this putative ability to memorise sentence fragments does not surface in children's production (Pinker 1984). Condition (ii), on the other hand, faces a problem with regard to its psychological feasibility, given that the number of possible human languages is potentially very large<sup>20</sup>. This has led to the postulation of additional conditions such as parameter ordering (see, for example, Dresher and Kaye 1990), the Independence Principle (Wexler and Manzini 1987), and hierarchical arrangement (Roeper and Weissenborn 1990).

However, it is important to note that these were developed in concordance with earlier formulations of the notion of parametric variation and it is by no means clear how they can be translated in more modern terms (and, more importantly, in terms that are compatible with our current background assumptions). Nevertheless, even if the Subset Principle - as formulated in (16) - turns out to be unusable (a conclusion suggested by Fodor and Sakas 2005) the conceptual point it makes is a valid one. An inductive learning mechanism that relies solely on positive evidence must be sufficiently conservative to prevent massive overshooting of the target.

The additional point that seems to emerge once we assume that language variation resides in the lexicon is that the language system must face at least two separate subset problems. On the one hand, the acquisition device must associate some properties/features with a given lexical item based on the item's linguistic behaviour. This is what we might call the 'classical' subset problem, as it corresponds to the type of scenario that Wexler and Manzini (1987) discussed in their original presentation of the Subset Principle. However, once the child has decided what features/properties to associate with a specific item, s/he must also decide what other lexical items these properties might be generalised to (cf. section 2.4.1). This latter process, which we might call the 'lexical' subset problem, appears to be what Fodor and Sakas (2005) have in mind at the beginning of their discussion on the role of the Subset Principle. Nevertheless, they do not explicitly consider the possibility that the subset problem could involve more than one

subpart. In fact, as far as I am aware, the possibility of such conceptual split has hitherto never been discussed<sup>21</sup>.

In order to illustrate the existence of two separate processes (i.e. a 'classical' and a 'lexical' problem) I will briefly examine an example in relation to the acquisition of anaphora (thus following a long tradition begun by Wexler and Manzini 1987 and recently revived in Hale and Reiss 2003).

It is now well known that anaphoric elements such as *himself/herself* show cross-linguistic differences in relation to their binding requirements. While English anaphors must be bound within the same clause, their Icelandic counterparts can depend on a higher antecedent:

- |     |  |           |
|-----|--|-----------|
| 17. | a. John <sub>i</sub> asked Peter <sub>j</sub> to shave himself <sub>*i/j</sub> | English   |
|     | b. Jón <sub>i</sub> skipaði Pétri <sub>j</sub> að raka sig <sub>i/j</sub>      | Icelandic |
|     | John ordered Peter to shave himself  |           |

The 'classical' subset question is the following. What is a learner to do when exposed to a sentence such as (18)? (Assuming that the learner understands the interpretation intended by the speaker).

18. John hurt himself

It seems that the learner must be conservative in that, given (18), s/he should assume that *himself* must be bound locally (i.e. s/he is being exposed to language (17a)). If it turns out that s/he is actually acquiring Icelandic, there will be positive evidence to indicate this, in the form of sentences such as (17b) in its non-local interpretation (i.e. *John<sub>i</sub> asked Peter<sub>j</sub> to shave himself<sub>i</sub>*). On the other hand, if s/he erroneously assumed that the target language allows non-local binding, there will be no amount of positive evidence that can signal the need for retreat. In other words, when the learner postulates a (bunch of) grammatical property(-ies) based on some input, s/he must be sufficiently conservative as to avoid fatal overshooting of the target. Let us assume - following a suggestion made by Hale and Reiss (2003) - that this is equivalent to associating the features [+bound] and [+local], as opposed to the single feature [+bound], with the element *himself*<sup>22</sup>. Although this process enables the child to take a desirably conservative step towards completing the acquisition of the lexical item *himself*, it has nothing to say about how such knowledge can be generalised to other elements. Should the child generalise the [+bound, +local] specification only to other instances of the same



anaphor, to all anaphors, to all referentially dependent elements, or even to all NPs?<sup>23</sup> This step too must involve a process sufficiently conservative as to avoid too wide a generalisation.

If the architecture proposed so far is on the right track, the Cascade-updating mechanism should offer an answer to this 'lexical' subset problem. First, the child will learn – via some conservative process – that some elements, say *himself* and *herself*, bear the properties [+bound, +local]. In turn, this will trigger application of Paradigm Formation and consequent identification of the set of properties that is being targeted. Then, and only then, will generalisation take place, thus affecting a whole paradigm.

It is important to note that this procedure is not conceptually equivalent to what I have been calling the 'classical' problem. Firstly, the 'classical' subset problem is not – strictly speaking – a problem of generalisation at all since it refers to the ability to ascribe a structure to a given input. Although this must necessarily be done in a conservative manner, so as to enable learning via positive evidence only, the process involved is one of interpreting rather than generalising. Moreover, this interpreting process seems to necessarily involve a mechanism that has access to grammatical principles. For example, a learning mechanism that on hearing the sentence in (18) decides to associate certain specific features with the anaphor must have knowledge of the principles that underlie such concepts as syntactic dependency and binding domains. It must also have access to structural information, such as syntactic configurations, as well as to the constraints applicable to them, such as minimality conditions, economy considerations, and obligatoriness requirements, to name but a few. Translated into the framework adopted here, the 'classical' problem requires a mechanism that has access to the inner workings of the linguistic modules as well as to the mapping principles. This differs sharply from the mechanism that deals with the other subset problem which – as I suggest in section 2.4.3 – acts on feature bundles, thus only needing access to lexical mappings. Although these two mechanisms are perhaps conceptually similar in that they must both function as to avoid overshooting of the target, it seems that they must necessarily be distinct in their formal characteristics.

Behind the formulation of Paradigm Formation is the idea of providing a formalisation for one of these mechanisms, namely the one which enables the acquisition device to go from item acquisition to paradigm acquisition. Ultimately, the major challenge we are faced with is that of developing a lexical model which can generalise in a somewhat conservative manner.

Naturally, this does not solve the 'classical' problem. However, it appears that we have made two important steps forward. Firstly, we have uncovered a hitherto undetected distinction between feature/property postulation on the one hand and lexical generalisation on the other. Moreover, we now have a formal definition of a potential solution to the 'lexical' subset problem. I will now consider how the definition of Cascade-updating may be further refined before moving on to investigate how its application could lead to the emergence of overgeneralisation as well as transfer effects.

#### 2.4.5 UG-Constrained Updating

Some of Chomsky's most compelling arguments in favour of the postulation of a language faculty involve considering linguistic knowledge that could not possibly be achieved by means of an inductive system like Cascade-updating. As is well known, children do not generalise just *any* property. Binding Principles are a case in point. On hearing a sentence like *Charlie Brown said he likes apples* a child might generalise that the pronoun *he* can refer to proper names and thus produce the structure in (19).

19. He<sub>i</sub> said that Charlie Brown<sub>i</sub> likes apples.

Nevertheless, it has often been reported that children do not make such errors (cf. Crain and Thornton 1998). Moreover, as has been shown by Gold (1967), a mechanism that relies solely on an inferential procedure would need more than a human lifetime to achieve the target language. The conclusion that has been drawn from this type of evidence is the same as that for which Chomsky has forcefully and convincingly argued in much of his work (e.g. Chomsky 1965, 1986b): the language acquisition device must be structured in such a way that it entertains only a small set of the logically possible hypotheses potentially compatible with the input.

A natural way of incorporating this insight into the Cascade-updating model is to limit its application in such a way that it stays within the bounds of Universal Grammar. In other words, to assume that the lexical mappings that Cascade-updating creates must be UG-compatible. Let us consider the acquisition of auxiliaries as an example. It is well known that children produce lexical content verbs before functional elements such as auxiliaries (see for example Radford 1990, 1996; van Kampen 2005). Given our background assumptions, this early stage will involve the child's lexicon hosting a number of elements mapped onto the

specification [+V, -N]. Among these, there may be some auxiliaries which the child might not yet have recognised as such<sup>24</sup>:

20.

- |    |                                      |   |                               |   |         |
|----|--------------------------------------|---|-------------------------------|---|---------|
| a. |                                      | ↔ | [+V, -N]                      | ↔ | /kæn/   |
| b. |                                      | ↔ | [+V, -N]                      | ↔ | /wɪl/   |
| c. | NP <sub>1</sub> _ (PP <sub>2</sub> ) | ↔ | [+V, -N]                      | ↔ | /spi:k/ |
|    |                                      |   | θ <sub>1</sub> θ <sub>2</sub> |   |         |

Given Paradigm Formation, the items in (20a-c) will be perceived as paradigmatically related in virtue of their sharing the [+V, -N] features. Consider now a later stage where the child comes to associate the feature [+A] with the verbs in (20b) and (20c) via the item-by-item procedure discussed in section 2.4.1:

21.

- |    |              |   |       |
|----|--------------|---|-------|
| a. | [+V, -N, +A] | ↔ | /kæn/ |
| b. | [+V, -N, +A] | ↔ | /wɪl/ |

Once this operation has taken place, Paradigm Formation will perceive the [+A] feature as targeting the whole [+V, -N] paradigm, thus triggering Cascade-updating with the following effect:

22.

- |    |                                      |   |                               |   |         |
|----|--------------------------------------|---|-------------------------------|---|---------|
| a. |                                      | ↔ | [+V, -N, +A]                  | ↔ | /kæn/   |
| b. |                                      | ↔ | [+V, -N, +A]                  | ↔ | /wɪl/   |
|    |                                      |   | ↓                             |   |         |
| c. | NP <sub>1</sub> _ (PP <sub>2</sub> ) | ↔ | [+V, -N, +A]                  | ↔ | /spi:k/ |
|    |                                      |   | θ <sub>1</sub> θ <sub>2</sub> |   |         |

This outcome, however, is undesirable on both conceptual as well as empirical grounds. Firstly, a generalisation of the kind exemplified in (22) involves massive

overshooting of the target, given that the number of lexical verbs contained in the child's lexicon will necessarily be much greater than the number of auxiliaries. If Cascade-updating were to lead to such a result it would inevitably cause a potentially irreversible state of affairs. Moreover, although overgeneralisation is common in child language, there are no reported examples of it applying on such a scale whereby a major lexical class is erroneously affected in the way that the example in (22) implies. This is exactly where the notion of UG-compatibility becomes important. Although a logical possibility, the mapping in (22c) is incompatible with UG as it involves associating a [+A] feature with an element that is already associated with an argument structure. However, auxiliaries – virtually by definition – lack a thematic/argument structure (see for example Pollock 1989, Poletto and Benincá 2004). Note that some kind of argument structure (albeit arguably underdeveloped) must be assumed for lexical verbs even at a relatively early stage since children do make distinctions between those verbs that take an object and those that do not<sup>25</sup>. Consequently, the putative outcome in (22c) is not only undesirable but also inadequate given universal considerations on what constitutes a [+V, -N, +A] element. We should therefore amend the Cascade-updating algorithm so as to incorporate UG-compatibility:

23. *Cascade Updating*

A property  $p$  that is newly acquired through the input items  $X, \dots, X^n$  will be attributed to item  $Y$  iff

- (i)  $Y$  possesses the paradigmatic properties common to  $X, \dots, X^n$  at the time when  $p$  is acquired and
- (ii)  $Y$  does not possess a value which is in conflict with  $p$ .

The clause in (ii) prevents Cascade-updating from adding the [+A] feature to any verb that possesses an argument structure, thus avoiding the undesirable operation depicted in (22).

Now that I have defined the Cascade-updating model in some detail, I will look at how its application may help us understand some instances of overgeneralisation that have been reported in the literature on child language.

### **3. Updating and Overgeneralisation**

#### **3.1 Introduction**

In the previous sections I claimed that the Cascade-updating model may offer a plausible answer to the 'lexical' subset problem. In this section I will discuss how its application, though principally concerned with productive generalisation, might occasionally land the learner in a superset situation. Note that this outcome is conceptually incompatible with the Subset Principle which, by definition, can only select a subset language. Arguably, however, a theory of language acquisition that bars overgeneralisation altogether is descriptively inadequate. As already mentioned above, overgeneralisation is a relatively common phenomenon in child language, a fact that must be taken into account by anyone who attempts to develop a successful model of language acquisition. In this section I will show that Cascade-updating does allow for overgeneralisation precisely in those domains where it has been reported to occur. Before this can be done, however, I will need to discuss some background assumptions with regard to the representation of predicate structure as well as the initial state of the language faculty.

#### **3.2 Acquisition and Innateness**

Chomsky (1986b) argues that the basic problem for any theory of UG is to construct an inventory of principles and properties which appropriately constrain the number of possible hypotheses to be entertained by the learner. To a certain extent, the Cascade-updating model is in itself a contribution towards such a theory for it provides a specific mechanism which is claimed to assist the learner in forming conservative hypotheses about the target language. However, as we have already seen, Cascade-updating can achieve descriptive adequacy only if it is embedded within a specific set of assumptions with regard to UG (cf. section 2.4.5). This means that, among other things, we must concern ourselves with considerations about the initial state of the language faculty. In this section I will discuss some necessary assumptions with regard to this issue.

##### *3.2.1 The Initial State*

It is commonly assumed that the initial state involves knowledge of the features that make up word categories<sup>26</sup> as well as the status of these categories within the syntactic component. If this is correct, features such as [N] or [V] will not need to be acquired, nor will the fact that they may combine. Instead, the child will only need to acquire which combination(s) of these features are part of the target and which specific properties map onto them.

The initial state need not involve only syntactic knowledge, however. Hale and Keyser (1996) argued that the semantic structure of predicates is based on three primitive conceptual relations, namely *states*, *events* and *interrelations*, represented by the features *s*, *e* and *r* respectively (see also Reinhart 2000). Predicates can be either monadic or dyadic with the exception of interrelations which, by definition, are only dyadic. A particular semantic implementation of this view has been proposed by Neeleman and van de Koot (2002:6), here presented as (24):

- |     |    |                                  |                           |
|-----|----|----------------------------------|---------------------------|
| 24. | a. | $\lambda x [{}_e x]$             | laugh, shake, work        |
|     | b. | $\lambda x \lambda y [{}_e x y]$ | kiss, hit, stroke         |
|     | c. | $\lambda x [{}_s x]$             | exist, green, intelligent |
|     | d. | $\lambda x \lambda y [{}_s x y]$ | know, see, understand     |
|     | e. | $\lambda x \lambda y [{}_r x y]$ | contain, have             |

The relation between a lambda operator and a variable is one of binding and it indicates that the variable is an argument of the semantic function to its right (this dates back to the work of Montague (1973)). The effect of the function is such that its arguments are associated with an elementary meaning, represented by the feature in the subscript. Thus, the example in (24a) denotes an event involving *x*, e.g. the event of 'working'. The structure in (24b) also denotes an event although in this case there are two entities involved, namely *x* and *y*. Similarly, the structures in (24c) and (24d) denote a state but while the former involves only one entity, the latter involves two. Finally, (24e) exemplifies the representation of an interrelation between two entities.

More complex structures can be derived via embedding. However, the possibilities are rather limited. Neeleman and van de Koot (2002) recognise only two such types of embedding, licensed by the following meaning postulates:

- |     |    |  |
|-----|----|--|
| 25. | a. | $[{}_e x [{}_{s/r} \dots y \dots]]$ , <i>x</i> affects <i>y</i> with the result that <i>y</i> obtains the property expressed by <i>s/r</i> . |
|     | b. | In $[{}_{s1} x [{}_{s2} \dots y \dots]]$ , <i>x</i> experiences <i>y</i> as having the property expressed by <i>s2</i> .                     |

However many predicates a language may have, they all result from the simple structures outlined in (24) or their combination with the two licensing conditions in (25). In other words, predicate structures, like word categories, are limited to a

handful of possibilities which are part of the child's knowledge from the initial state. Unarguably, this pre-existing knowledge is invaluable to the child as it provides accurate guidelines with regard to what can be expected to occur, essentially the difference between general learning and language acquisition.

Such guidelines are not confined to the acquisition of properties (cf. (9a). As Jackendoff (1997) points out, correspondence rules do not allow just any type of mapping as they are subject to universal constraints which limit the possible mapping relations that can be established. From an acquisitional perspective, these constraints enhance learnability by limiting the idiosyncrasies that are expected to arise, thus greatly simplifying the task in (9b).

### 3.3 Mapping Principles

An example of a universal (or general) constraint is that which governs the mapping of a semantic argument structure onto theta-roles (adapted from Neeleman and van de Koot, 2002):

26. Each lambda operator, its restriction and the variable(s) it binds corresponds to a  $\theta$ -role and vice versa.

This principle, which is taken to hold universally, will also be part of the initial state thus providing the child with the knowledge that while (27a) is a possible mapping structure for a lexical item, (27b) and (27c) are not:

27.

- |    |                                    |                   |                                |                   |          |
|----|------------------------------------|-------------------|--------------------------------|-------------------|----------|
| a. | $\lambda x [{}_e x]$               | $\leftrightarrow$ | $[+V, -N], \theta_x$           | $\leftrightarrow$ | $/la:f/$ |
| b. | * $\lambda x [{}_e x]$             | $\leftrightarrow$ | $[+V, -N], \theta_x, \theta_y$ | $\leftrightarrow$ | $/la:f/$ |
| c. | * $\lambda y \lambda x [{}_e x y]$ | $\leftrightarrow$ | $[+V, -N], \theta_x$           | $\leftrightarrow$ | $/hit/$  |

The rule in (26) expresses the architectural relation that holds between the conceptual component and syntax. Presumably, the syntactic component evolved as a means for structuring complex conceptual relations which could not be expressed linearly (Bickerton 1990, Jackendoff 1997). If this is the case, it is expected that its structure is affected by that of the conceptual system. In this sense, the term "constraint" is actually inaccurate, as we are not dealing with a restriction imposed in order to limit the structure of theta-grids. Quite the contrary, theta-grids can only

have the structure they do because they developed out of conceptual argument structures that underlie the specific manner in which concepts can relate or, perhaps more precisely, the manner in which humans perceive relations between concepts.

### 3.3.1 Variable Binding and Syntax/Semantics Mapping

Although the principle in (26) states that there must be a one-to-one mapping between lambda bound variables and thematic roles, it does not follow that variables must necessarily be bound by a lambda operator. Assuming that variables must be bound in order to yield an interpretation, variables that are not lambda-bound can still enter a binding relation by undergoing existential closure. Following the mapping principle in (26), existentially bound variables will not be mapped onto theta roles, thus creating a mismatch between the number of semantic and syntactic arguments. An example of this mismatch is found in those predicates that are optionally transitive like *eat*. A structure such as 'John is eating' implies that there exists something that John eats. This implication arises as a result of the internal argument being existentially-bound as opposed to lambda-bound:

28.

$$\exists y \lambda x [{}_e x y] \quad \leftrightarrow \quad [+V, -N], \theta_x \quad \leftrightarrow \quad /i:t/$$

In the case of *eat*, existential closure is a possibility but so is binding by a lambda operator. The complete set of properties for *eat* will therefore include both mappings:

29.

$$\begin{array}{llll} \exists y \lambda x [{}_e x y] & \leftrightarrow & [+V, -N], \theta_x & \leftrightarrow & /i:t/ \\ \lambda y \lambda x [{}_e x y] & \leftrightarrow & \theta_x \theta_y & & \end{array}$$

Note that, unlike traditional lexicalist models (e.g. Jackendoff 1975, Pinker 1984), representational modularity does not raise the question of whether verbs like *eat* have multiple entries. This is because – within this model – lexical structures are not entries as such but spell out rules which specify how a bundle of features/properties is mapped onto different components and, ultimately, phonologically realised (see also mapping in (62) below). The representation in (29) can therefore be understood as the grouping of two different feature bundles mapped onto the same phonological string<sup>27</sup>.



However, not all predicates allow both these options. A well-known minimal pair is given by *eat* and *devour* (Jackendoff 1990). While the former allows the syntax-semantic mapping in (28), the latter does not as both its arguments are obligatorily lambda-bound:

30. a. \*The tyrannosaurus devoured  
b. The tyrannosaurus devoured its prey

In the same vein, there are predicates that require some argument(s) to be obligatorily bound by an existential, thus disallowing lambda-binding altogether. The minimal pair *send* and *drop* illustrates this:

31. a. Bill sent a present  
a'. Bill sent John a present  
  
b. Bill dropped a stone  
b'. \*Bill dropped his foot a stone

Like *send*, *drop* is an event that embeds an interrelation. If '*Bill drops a stone*' he initiates an event that results in the stone entering an interrelation with a locus. Nevertheless, as (31b') shows, the argument expressing the locus disallows lambda binding. The lexical properties of the two predicates will therefore differ with respect to the binding relations they allow and the consequent thematic mapping:

- 32.
- |   |                   |                               |                   |               |
|---|-------------------|-------------------------------|-------------------|---------------|
| $\lambda z \exists y \lambda x [{}_e x [{}_r y z]]$ | $\leftrightarrow$ | $[+V, -N], \theta_x \theta_z$ | $\leftrightarrow$ | <i>/send/</i> |
| $\lambda z \lambda y \lambda x [{}_e x [{}_r y z]]$ | $\leftrightarrow$ | $\theta_x \theta_y \theta_z$  |                   |               |
|   |                   |                               |                   |               |
| $\lambda z \exists y \lambda x [{}_e x [{}_r y z]]$ | $\leftrightarrow$ | $[+V, -N], \theta_x \theta_z$ | $\leftrightarrow$ | <i>/drop/</i> |

In sum, the conceptual system imposes no limitation on variable binding and any argument could, in principle, be either lambda-bound or existentially-bound. Nevertheless, predicates vary with regard to whether or not they allow both relations. These variations are language-specific and need to be acquired as part of the target<sup>28</sup>. This fact, in conjunction with the Cascade-updating mechanism, may lead to certain types of overgeneralisation.

### 3.4 The Transitive/Unaccusative Alternation

As is well known, there is a rather large set of verbs in English that display a transitive/unaccusative alternation (also known as causative/inchoative alternation). These are verbs such as *break* and *open*:

33. a. The door opened  
a'. The vase broke
- b. John opened the door  
b'. Bill broke the vase

The semantic structure of these verbs involves embedding of a *state* in an *event* which, by virtue of the meaning postulate in (25), will be associated with an interpretation whereby the *state* results from the *event*. It is clear that they are not simple event structures since they cannot denote unbounded activities, as shown by their inability to take a duration phrase (see Jackendoff 1990 for details):

34. a. \*The door opened for hours  
a'. \*The vase broke for hours
- b. \*John opened the door for hours (only repetition)  
b'. \*Bill broke the vase for hours

The alternation between the examples in (33a) and (33b) shows that, parallel to (29) and (31a-a'), *open* and *break* allow existential as well as lambda binding of one of the argument variables. However, contrary to the predicates discussed in section 3.3.1, it is the most prominent argument variable that may undergo either type of binding. This results in the following mapping relations:

- 35.
- |  |                   |                      |                   |         |
|--|-------------------|----------------------|-------------------|---------|
| $\lambda y \exists x [e \times [s \ y]]$ | $\leftrightarrow$ | $[+V, -N], \theta_y$ | $\leftrightarrow$ | /breɪk/ |
| $\lambda y \lambda x [e \times [s \ y]]$ | $\leftrightarrow$ | $\theta_x \theta_y$  |                   |         |
|  |                   |                      |                   |         |
| $\lambda y \exists x [e \times [s \ y]]$ | $\leftrightarrow$ | $[+V, -N], \theta_y$ | $\leftrightarrow$ | /əʊpən/ |
| $\lambda y \lambda x [e \times [s \ y]]$ | $\leftrightarrow$ | $\theta_x \theta_y$  |                   |         |

By undergoing existential binding, the most prominent argument is not mapped onto a thematic role, thus resulting in the archetypal unaccusative structure that features an embedded argument as the syntactic subject. This alternation is not dependent on the type of embedded structure as it is also found in predicates that embed an interrelation (typically, but not exclusively, between an entity and a locus), such as *trip* and *sink*:

36. a. John tripped  
a'. The boat sank  
  
b. Mary tripped John  
b'. Susan sank the boat

The examples in (36) are parallel to those in (33) as the a-type structures differ from the b-type structures with regard to the operator that binds the argument associated with the event. Their mapping relations will therefore pattern with those in (35), except for the type of structure they embed:

- 37.
- |  |                   |                      |                   |        |
|--|-------------------|----------------------|-------------------|--------|
| $\exists z \lambda y \exists x [e \times [r, y, z]]$ | $\leftrightarrow$ | $[+V, -N], \theta_y$ | $\leftrightarrow$ | /trip/ |
| $\exists z \lambda y \lambda x [e \times [r, y, z]]$ | $\leftrightarrow$ | $\theta_x \theta_y$  |                   |        |
| $\exists z \lambda y \exists x [e \times [r, y, z]]$ | $\leftrightarrow$ | $[+V, -N], \theta_y$ | $\leftrightarrow$ | /sink/ |
| $\exists z \lambda y \lambda x [e \times [r, y, z]]$ | $\leftrightarrow$ | $\theta_x \theta_y$  |                   |        |

Given the definition of Cascade-updating, on exposure to a small set of verbs with structures like those in (35) and (37), the language system will take complex events to form a paradigm with the following as its paradigmatic properties<sup>29</sup>:

38. ..  $\exists x$  ..  $[e \times [...]]$       [+V, -N]

Any lexical item that is associated with these properties will therefore be affected by the application of Cascade-updating, thus resulting in a large number of verbs acquiring the alternative lambda-bound structure as follows:

39.

$\lambda y \exists x [{}_e x [{}_s y]]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/breik/
$\lambda y \lambda x [{}_e x [{}_s y]]$	$\leftrightarrow$	$\theta_x \theta_y$		
$\lambda y \exists x [{}_e x [{}_s y]]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/ʃʌt/
$\lambda y \lambda x [{}_e x [{}_s y]]$	$\leftrightarrow$	$\theta_x \theta_y$		

As a result of the process in (39) the system will have acquired the properties necessary in order to make the distinction between *John shut the door* and *the door shut* without having to be exposed to any tokens of the former.

Since the number of verbs that present the unergative/transitive alternation is rather large, it would take a considerable amount of time should the relevant lambda-bound structure have to be acquired piecemeal. Therefore, the application of Cascade-updating to this paradigm constitutes a major advantage to the learner.

### 3.5 The Nature of Overgeneralization

As shown by the examples in (30) and (31), there is some arbitrariness as to which verbs allow both types of binding. Nevertheless, Cascade-updating applies indiscriminately affecting all those items that share the relevant paradigmatic properties. This predicts that Cascade-updating will inappropriately affect those items which share the properties in (38) but so happen to arbitrarily disallow binding by a lambda operator.

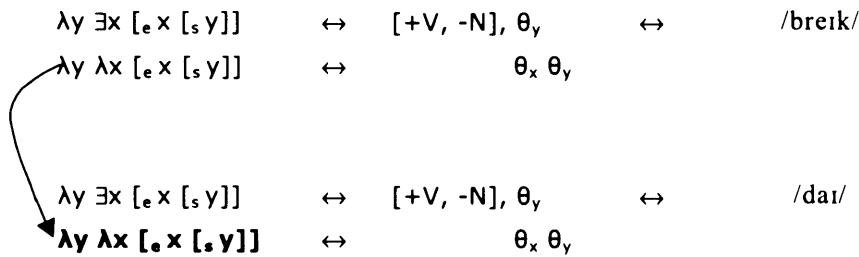
The number of predicates that display this arbitrariness is rather small when it comes to events embedding a state. In fact, it is extremely small when we consider those predicates that could realistically be part of a child's lexicon. One such verb is *die* which, unlike *break* and *open*, requires its most prominent argument to be obligatorily existentially bound:

40.

$\lambda y \exists x [{}_e x [{}_s y]]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/daɪ/
---	-------------------	----------------------	-------------------	-------

Nevertheless, due to its sharing of the paradigmatic properties that triggered Cascade-updating in (39), the entry in (40) will acquire the alternative lambda structure:

41.



This process underlies some well-known instances of overgeneralisation, such as the following:

42. (From Bowerman 1982:14-15)

- |  |             |
|--|-------------|
| a. Who <i>deaded</i> my kitty cat?                   | John 2;6    |
| b. He's gonna <i>die</i> you.                        | Hilary 4+   |
| c. Don't <i>dead</i> him.                            | Eva 4;10    |
| d. If you want it to die Eva is gonna <i>die</i> it. | Christy 5;0 |

A much larger number of cases that arbitrarily exclude lambda-binding is exemplified by events that embed interrelations. Verbs of this kind include *come*, *go*, *fall*, *disappear* and *rise*. These are also expected to be affected by Cascade-updating and thus to acquire the alternative lambda structure that results in the realisation of a second syntactic argument. As shown in (43), this is indeed the case (from Bowerman 1982; 13-15):

- |   |              |
|---|--------------|
| 43. a. I <i>come</i> it closer so it won't fall.  | Christy 2;9  |
| b. She <i>came</i> it over here.  | Christy 3;4  |
| 44. a. Go me to the bathroom before you go to bed.  | Christy 3;10 |
| b. Do you have anything else you'd like to go to China?<br>(pretending to drive a train carrying freight to China). | Eva 4;11     |
| 45. a. I'm gonna just <i>fall</i> this on her.  | Christy 2;9  |
| b. And the doggie had a head and somebody <i>fell</i> it off.   | Eva 3;8      |
| 46. a. Why do we have to <i>rise</i> it?  | Marcy 6;4    |
| b. It's <i>rising</i> me.   | Christy 6;8  |

47. a. I'm gonna put the washrag in and *disappear* something    Eva 3;7  
under the washrag.  
b. I *disappeared* a bear in the back of the car.    Scott 5;0

As expected, presence of the paradigmatic properties in (38) triggers the application of Cascade-updating, giving rise to the structures underlying (42) to (47). Although the effects represented in (39) and (41) are identical, while the former takes a step further towards the target language, the latter gives rise to an overgeneralization which will have to be rectified. The very process that makes the language system efficient, then, is also responsible for the occurrence of overgeneralization.

Note that most of the lambda structures realised in (42) to (47) are indeed possible in the target language. However, it so happens that they are realised as separate mapping relations involving different phonological properties, namely /kill/, /bring/, /teik/, /drop/ and /reiz/ respectively for (42) to (46). These language specific mappings will presumably need to be acquired before the effects of Cascade-updating can be overridden (see also a brief discussion in section 4.1).

### 3.6 The Simplex/Complex Alternation

Another large paradigm found in English involves predicates that alternate between a simple *interrelation/event* and a complex *event* reading (interrelations of this type are usually locative):

48. a. The child stood for hours  
b. Mary stood the child on the table
49. a. Music played for hours  
b. John played music for hours

The semantic structure of the a-type predicates is that of a simple *interrelation* or *event*, while the b-type predicates are complex events which embed an interrelation, as in (48), or a further event<sup>30</sup>, as in (49). These structures will be lexically encoded as follows<sup>31</sup>:

50.

$\exists y \lambda x [{}_r x y]$	$\leftrightarrow$	$[+V, -N], \theta_x$	$\leftrightarrow$	/stænd/
$\exists y \lambda x \lambda z [{}_e z [{}_r x y]]$	$\leftrightarrow$	$\theta_z \theta_x$		
$\lambda x [{}_e x]$	$\leftrightarrow$	$[+V, -N], \theta_x$	$\leftrightarrow$	/plei/
$\lambda x \lambda y [{}_e y [{}_e x]]$	$\leftrightarrow$	$\theta_y \theta_x$		

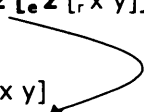
Predicates with the structure above will then be perceived to form a paradigm not very much different from the one in (38) (the alpha subscript symbolises the irrelevance of the predicate type):

51.    ..  $\lambda x$  ..  $[{}_o x]$                        $[+V, -N]$

In other words, all verbs that have a non-embedded structure will be taken to form a paradigm. Presence of the paradigmatic properties in (51) will then result in the now familiar application of Cascade-updating, with consequent acquisition of the alternative embedding structure:

52.

$\exists y \lambda x [{}_r x y]$	$\leftrightarrow$	$[+V, -N], \theta_y \theta_y$	$\leftrightarrow$	/stænd/
$\exists y \lambda x \lambda z [{}_e z [{}_r x y]]$		$\theta_x \theta_y \theta_z$		
$\exists y \lambda x [{}_r x y]$	$\leftrightarrow$	$[+V, -N], \theta_x \theta_y$	$\leftrightarrow$	/sit/
$\exists y \lambda x \lambda z [{}_e z [{}_r x y]]$		$\theta_x \theta_y \theta_z$		



Parallel to (39), the system will have acquired the properties necessary in order to make the distinction between *John sat for hours* and *Mary sat the baby on the stool* without having to be exposed to any tokens of the latter.

Similarly to the unaccusative/transitive case, however, there are verbs that do not display the alternation despite sharing the relevant paradigmatic properties. A clear example of this arbitrariness is provided by *sing* which, unlike *play*, does not result from embedding: *I sang a song*, *\*the song was singing*. Examples of such verbs are the following<sup>32</sup>:

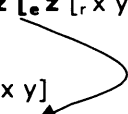
53.

$\lambda y \exists x [{}_r x y]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/steɪ/
$\lambda y \lambda x [{}_s x y]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/ɾɪmembə/
$\lambda x [{}_s x]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/bi:/
$\lambda x [{}_e x]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/gɪgəl/

Nevertheless, the fact that they share the relevant paradigmatic properties will trigger Cascade-updating, thus resulting in the acquisition of the alternative lambda structure:

54.

$\lambda y \exists x [{}_r x y]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/stænd/
$\lambda y \exists x \lambda z [{}_e z [{}_r x y]]$				
$\lambda y \exists x [{}_r x y]$	$\leftrightarrow$	$[+V, -N], \theta_y$	$\leftrightarrow$	/steɪ/
$\lambda y \exists x \lambda z [{}_e z [{}_r x y]]$				



The data in (55) to (59) illustrate how these overgeneralisations surface in the speech of English children:

55. (From Bowerman 1982:14-15)

- |   |             |
|---|-------------|
| a. I'm <i>staying</i> it in the water.        | Eva 3;2     |
| b. I want to <i>stay</i> this rubber band on. | Christy 3;7 |

- |   |           |
|---|-----------|
| 56. a. I keep have to <i>remember</i> you<br>(complaining about M forgetfulness). | Eva 4;11  |
| b. I have to <i>remember</i> my daddy Saturday<br>Winnie the Pooh is on.          | Mindy 6;7 |

- |   |             |
|---|-------------|
| 57. a. I want to <i>have</i> it to Mark.      | Christy 4;2 |
| b. Would you like me to <i>have</i> you some? | Christy 4;6 |

- |  |             |
|--|-------------|
| 58. <i>Be</i> a picture of Emily and me. | Christy 3;5 |
|--|-------------|

- |   |            |
|---|------------|
| 59. Do you want to come watch the mans <i>sing</i> their guitars? | Emily 2;11 |
|---|------------|



Similarly to the examples in (43)-(47), these data show how detection of the relevant paradigmatic properties triggers the application of Cascade-updating, giving rise to structures that diverge from the target.

As noted for examples (42) to (47), the structures underlying (55) to (59) are also possible in the target language, although they are realised as separate mapping relations involving different phonological properties.

### 3.7 Alternative Analysis: morphological affixation

The architecture outlined so far relies on the idea that alternations such as transitive/unaccusative involve lexical items with more than one semantic structure. An alternative view would be to postulate only one semantic structure for each predicate and to attribute the alternation to the application of a phonologically empty affix. Within the framework adopted here, such affix would consist of a set of semantic properties realised as an independent mapping in the form of the following template:

60.    ..  $\lambda x [{}_e x \text{ \_\_\_ }]$

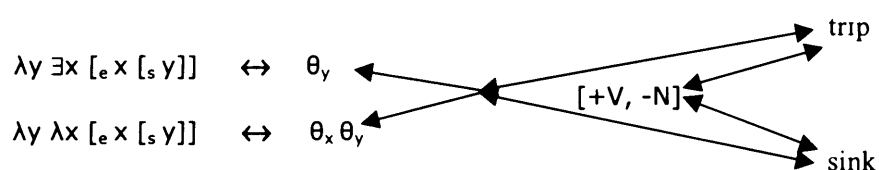
The postulation of the item in (60), however, has a number of shortcomings. Firstly, it is unclear what syntactic properties it would have besides the *ad hoc* feature [affix]. Moreover, it is by no means obvious how such an affix could be acquired. Alternations such as *the vase broke* and *John broke the vase* provide unambiguous cues with regard to the presence of more than one semantic structure for *break*. However, a set of mapping relations must – by definition – involve properties from more than one component. Crucially, there are neither phonological nor syntactic differences between the two uses of *break*, thus excluding the postulation of a separate mapping such as (60)<sup>33</sup>.

Moreover, the only potential advantage of assuming a morphological process appears to be the elimination of storage since, contrary to the examples in (50), postulation of a separate affix such as (60) would allow the alternations to be morphologically derived, thus reducing lexical storage. Nevertheless, this assumption does not deliver the goods. As noted earlier, the distinctions are largely arbitrary:

61. The UFO broke/\*demolished/\*destroyed/disintegrated/shattered/\*wrecked

This would mean that many verbs should include lexical information about whether they allow the supposed morphological affixation, a rather unattractive characteristic for a process that would presumably avoid lexical storage. Both views must therefore resort to storage of lexical information in order to capture the arbitrariness of such alternations. Furthermore, the morphological view relies on the questionable acquisition of a further lexical item. It therefore seems that an analysis based on the lexicalization of semantic properties should be preferred. Moreover, it must be noted that, contrary to traditional views of the lexicon, mappings such as those in (37) do not involve multiple storage of the same piece of information. The role of cross-modular mappings is that of relating phonological information to semantic and syntactic properties. However, the advantage of a Representationally Modular approach to the lexicon is that these properties only need to be stored once, the only task that is repeated for each item is that of mapping them on more than one phonological string. For ease of presentation, I chose a notation that does not spell this out. However, mappings such as those in (37) would be more accurately depicted as in (62).

62.



In sum, not only does assuming a phonologically empty suffix raise learnability questions but it also brings no advantage in terms of storage when compared to a representationally modular model.

## 4. Retreating from a Non-target Lexicon

### 4.1 Introduction

Besides adding new information, language acquisition also involves the process of retreating from erroneously acquired information. The updating illustrated in (41) is an example of a decision to add extra information from which the language system will eventually have to backtrack. The question of what might persuade the language system to retreat is notoriously complex and one which, unfortunately, I do not have anything to contribute to. I will simply follow a suggestion by Pinker (1984) and assume that a lexical mapping is reinforced whenever the child's input provides evidence for it. As non-target structures that have been arrived at via

generalisation will never be reinforced, they will become increasingly weaker over time, and will eventually be phased out altogether (though, as I argue below, this must not result in the removal of lexical information).

What I would like to focus on in more details is how the backtracking process might function once the decision has been made. In particular, I will claim that actual retreat is undesirable and that if a system is to successfully acquire the target language it must function in an *informationally monotonic* manner.

#### 4.2 Deleting Information

After new information has been erroneously added, input will suggest that some postulated lexical properties are in fact divergent from the target language. In the case of (41) above a set of properties has been erroneously associated with the lexical item *die*. A system that allows deletion of lexical information will rectify this by removing the divergent properties, thus converting (41) - repeated here for convenience as (63a) - into (63b):

63.

- a.  $\lambda y \exists x [e \times [s \ y]] \leftrightarrow [+V, -N], \theta_y \leftrightarrow /dai/$   
 $\lambda y \lambda x [e \times [s \ y]] \theta_x, \theta_y$
- b.  $\lambda y \exists x [e \times [s \ y]] \leftrightarrow [+V, -N], \theta_y \leftrightarrow /dai/$

Within the model being proposed here, this operation has a disastrous outcome as the predicate has now reverted to the very structure that triggered Cascade-updating in the first instance. This would result in the system entering a loop whereby deletion will be followed by a further application of the updating process, which will then trigger further deletion and so on *at infinitum*.

This problem stems from the monadic nature of the process in (63) above, as discussed in section 2.1. As far as lexical features are concerned, we saw that the problem can be avoided by adopting a binary system. However, the same solution does not seem to be applicable to the domain of lexical properties, as it seems counterintuitive to assume that properties such as argument structures or theta grids are marked as + or -.

I propose to deal with this problem by postulating that the acquisition device operates in an informationally monotonic fashion. In other words, the acquisition device cannot remove any information from a lexical mapping; it can only add information. If so, 'deletion' of lexical information must be achieved by the addition

of information. Note that this view allows us to maintain the earlier claim that absence of a property entails lack of knowledge (see section 2.1). Absence of a piece of linguistic information therefore only occurs at the start of acquisition; it is a state to which the system cannot regress.

An informationally monotonic system is not unable to alter its decisions. It just cannot do so through deletion. Instead, it may mark the unavailability of a feature/property by recording that it is unavailable. This process will be indicated throughout by marking 'deleted' information with strike-through:

64.

$$\lambda y \exists x [e \times [s \ y]] \leftrightarrow [+V, -N], \theta_y \leftrightarrow /daɪ/$$

$$\cancel{\lambda x \lambda y [e \times [s \ y]]}$$

This process is consistent with monotonicity as it involves adding, rather than removing, information. Also, it must be noted that the piece of information being added is not of a linguistic nature, as the resulting lexical mapping is computationally equivalent to a mapping for which the argument structure at issue is absent. Marking a property as unavailable does not add any linguistically computable information; it simply keeps track of a decision to retreat from a certain setting (see Fodor and Sakas 2005 for extensive discussion on how successful acquisition is dependent on the presence of a memory device). This allows the system to mark the unavailability of a particular structure despite its inability to delete information.

This process is essentially a formalisation of the “phasing out” stage mentioned in section 4.1 above. The additional procedure – forced upon us by informational monotonicity – is that lexical information that is not reinforced by the input is eventually marked as unavailable instead of simply fading away. For this purpose, it may be useful to view the reinforcement process as an increase in the weight of a lexical connection, in a manner reminiscent to the weighing process employed in cognitive networks (see for example Hudson 2007, Luger and Stubblefield 1993). The strikethrough can then be seen as a sort of “threshold marker” which is applied to any lexical information that has never been reinforced, and has thus fallen below threshold.

The commitment to monotonicity also has an effect on the way we must view the treatment of lexical features. Switching between two separate feature settings, say from  $+f$  to  $-f$ , involves a process like the one in (65), which is similar to that in (64) above.

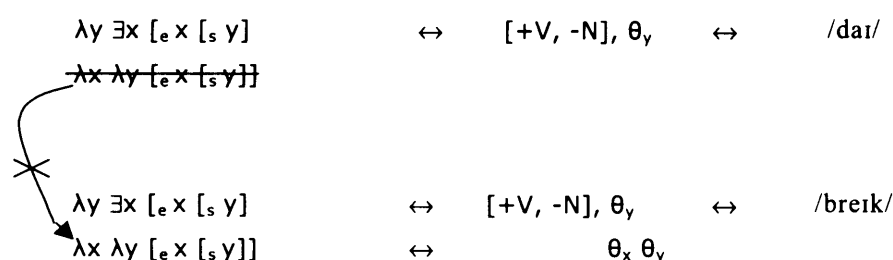
65.

wrong setting falls below threshold	+f
↓	
marking of unavailability	+f
↓	
addition of new setting	-f +f

Strictly speaking, the system does not actually retreat; it takes a further step towards the target by marking, and thus rejecting, a particular possibility.

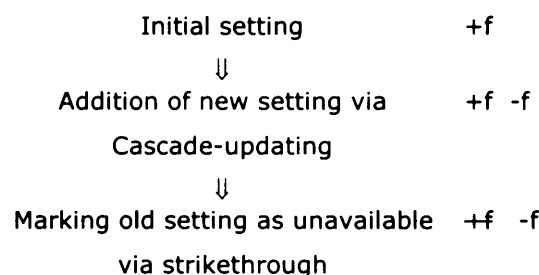
This architecture has one further implication: unavailability of a feature/property must be marked on an item-by-item basis, since Cascade-updating only deals with purely linguistic information. The strikethrough, being a memory device, does not encode any linguistic information and it is thus not a transferable property. Consequently, the system cannot perform en-bloc retreat. This is a welcome result as it prevents overgeneralisation of deletion, hence banning the process in (66).

66.



The same holds for lexical features. In cases where a specific feature, say  $+f$ , has been marked as unavailable for a lexical item  $X$  (step 1 in (65) above), unavailability will not be transferred to other items even if these may be paradigmatically related to  $X$ . Moreover, the fact that Cascade-updating is constrained by UG-compatibility (cf. section 2.4.5) also prevents deletion from spreading by indirect means. Consider a case where the process in (65) has taken place for some lexical item  $X$ . The newly acquired  $-f$  feature could then be generalised to other items paradigmatically related to  $X$ , a step which might then trigger subsequent application of the strikethrough, as in (67).

67.



However, for such a sequence to be completed successfully it is necessary that Cascade-updating adds a setting that is in conflict with a feature value already present in the lexical mapping at issue. This, which is represented by the intermediate step in (67), is ruled out by clause (ii) of Cascade-updating, as defined in (23) above. Consequently, the nature of Cascade-updating, together with the concept of conflicting properties, are partly responsible for restricting retreat to an item-by-item process.

In sum, informationally monotonicity forces the introduction of a memory device, indicated here by strikethrough, which prevents the language system from entering a deadlock whereby all items are being continually reset. Moreover, as a memory device, strikethrough can not be acquired via Cascade-updating, thus excluding the possibility of it being generalised. This suggests that the Cascade-updating model may be able to strike a balance between conservatism and positive generalisation, thus offering a potentially satisfactory model of language acquisition.

## 5. Interim Summary

In sections 2-4 of this chapter I have proposed a representation of lexical knowledge that is constituted by mapping relations between the semantic, syntactic and phonological components. I have then suggested how, within this representation, we can view some core properties of the language acquisition process.

I have shown how such representation is in tune with the ability to generalise which is characteristic of the acquisition process, as well as how this inevitably leads to some well-known overgeneralization effects. This was done by what I suggested is an informationally monotonic mechanism associated with the language system, namely *Cascade-updating* and its interaction with two primitive concepts: *i) paradigmatic properties ii) conflicting properties*. I then argued that

informational monotonicity is a necessary condition for the successful acquisition of the target language. On the one hand, Cascade-updating attributes new information to some items without necessarily having to receive them as input. On the other hand, if such items did occur in the input and have already been allocated a value for that particular property, Cascade-updating does not affect them.

In the next section I will investigate how the application of the very same processes can shed light on the effects of bilingual acquisition.

## **6. Bilingual Acquisition**

### **6.1 How Can a System Deal with Two Languages?**

The architecture presented so far is based on a view of language in which the components of grammar are universal, and cross-linguistic variation is determined by the occurrence of different lexical mappings, where all idiosyncrasies are reflected (Borer 1986, Chomsky 1995, Jackendoff 1997). There is a syntactic component that is identical for all languages and there are lexical items with idiosyncratic properties that, when inserted into that component, behave in a language-specific manner depending on their properties. In other words, English and, say, Japanese, are just the results of two different sets of items being put through the same universal system.

By assuming that the lexicon is central to language variation, it seems unavoidable that in order to understand bilingualism we need to take a look at the organisation of the bilingual lexicon.

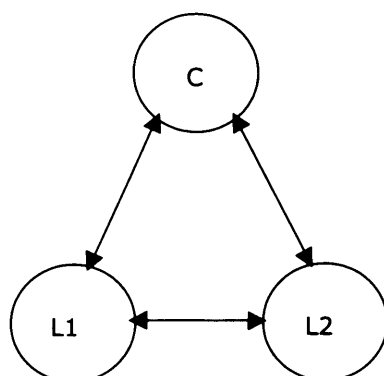
#### *6.1.1 The Bilingual Lexicon*

The question of whether bilinguals store items within one or two lexicons has long been object of debate within the acquisitional as well as the L2 literature. Although there have been various proposals as to the extent to which bilinguals integrate the two languages, it is generally agreed that lexical information in bilinguals does not involve total separation of the languages. Overwhelming evidence against separate lexicons has been collected through a variety of experimental techniques, which include semantic priming (Chen & Ng 1989, Schwanenflugel & Rey 1986, Keatley et al. 1994), cross-language categorisation (Caramazza & Brones 1980), and – more recently – neuro-imaging studies (Chee, Tan, & Thiel 1999, Illes et al. 1999, Klein, Milner et al 1995). This literature points in the direction of a mixed lexicon where representations are connected directly as well as indirectly through the conceptual system. Among other things, experimental evidence suggests that the amount of

representational properties that words from the two languages share is dependent on the particular type of word. For example, words referring to concrete entities as well as cognates (i.e. near homophones) have been found to pattern differently from other types of words. This cannot be accommodated within a model that assumes only indirect connections via the conceptual store, since this would not permit connections at the phonological or semantic level.

This led to the development of mixed hierarchical models of the type in (68), where C represents the conceptual store while L1 and L2 represent two separate lexical stores (from De Groot (1993), see also De Groot et al. (1994) for a network implementation).

68.



If the model in (68) is correct, we might expect that Paradigm Formation would identify separate paradigmatic relations for each of the languages, as it would presumably be unable to 'see' both languages simultaneously.

However, recent work on bilingual memory organisation, especially – though not exclusively – within the connectionist modelling tradition, has shown that language separation need not involve separate lexicons. In particular, French (1998) and Li and Farkas (2002) have argued for a view of the bilingual lexicon which involves one single storage. Within these models the problem of lexical selection, exemplified by the fact that bilinguals have no trouble remaining within a single language or recognising a word as belonging to a specific language, is solved by the emergence of organisation. This is a process whereby words from the two languages are clustered separately based on their separate phonological regularities, despite the fact that they are stored in a single lexicon. Dijkstra and Van Heuven (2002) also proposed a model that relies on a single lexical store. In fact, it appears that the vast majority of researchers investigating the nature of



bilingual memory are now in favour of a single lexical store (see French and Jacquet (2004) for a brief review).

This suggests that a bilingual lexicon is far from being equivalent to two monolingual ones. Rather, it may be more appropriate to view it as analogous to a monolingual lexicon which is organised into two lexical sets. Interestingly, this implies that the Separate Development Hypothesis (SDH) and the Single System Hypothesis (SSH) may both have been correct, albeit at different levels of observation. On the one hand, the SSH may have been correct to the extent that the two languages may indeed turn out to be stored together, but it is false in so far as it states that the child has no procedure which allows him/her to mark language separation. On the other hand, the SDH may be correct to the extent that it is committed to the existence of some process that keeps the two languages separate. However, this hypothesis may simply be describing a surface effect rather than referring to the underlying structure of lexical storage.

Most importantly, if it is indeed the case that lexical storage in bilinguals does not involve two separate lexicons, Paradigm Formation will enjoy a field of application which is much wider than what the architecture in (68) suggested. If the two lexical sets emerge as a consequence of phonological regularities, there may be no distinction at the semantic or syntactic level which could force Paradigm Formation to scrutinise one particular lexical set at a time. Consequently, the extraction of paradigmatic relations can, in principle, affect elements from both lexical sets. This process, I will argue, is what gives rise to transfer effects in bilinguals.

## **6.2 Transfer Effects as Overgeneralizations**

De Houwer (1990) proposes an informal definition of transfer effects as “items from one language [that] are structured as they would be in the other”. Since we have assumed that the behaviour of lexical items is dictated by the properties they comprise, for this situation to occur the items at issue must somehow acquire properties from their counterpart in the other L1.

Given the architecture of the acquisition device proposed in this chapter and the organisation of the bilingual lexicon discussed in section 6.1.2, it seems unsurprising that lexical items affect each other across languages as they do within the same language. When Paradigm Formation extracts paradigmatic relations, lexical items from both languages may be considered, so long as they host the relevant properties for the paradigm at issue. When this occurs, the consequent application of Cascade-updating will affect items from both languages, thus creating

what - in descriptive terms - have been known as syntactic transfer effects. In a sense, the traditional terminology is actually rather accurate. If the view presented here is correct, it is indeed the case that transfer effects involve a 'transferring' action whereby some feature/property is transferred from one lexical mapping to another. However, the term 'syntactic transfer' turns out to be inaccurate, since it is lexical features/properties that are subject to transfer, not syntactic structures. Throughout this thesis I will keep referring to the phenomenon as 'transfer effects', though I will occasionally use the term 'transfer' to refer to the underlying process, namely cross-language generalisation. This process is discussed in detail in chapter 2.

### 6.2.1 Paradigmatic Identity

For Cascade-updating to apply it is essential that the updated item meets two specific conditions: it must lack a *conflicting value* for the newly acquired property and it must possess the relevant *paradigmatic properties*. Consequently, the Cascade-updating model predicts that transfer effects will, like overgeneralization, apply to specific constructions. In chapters 3 and 4 I will discuss evidence in relation to this claim.

Another claim made by the Cascade-updating model regards informational monotonicity, and the constraints it places on retreat. This property too restricts the occurrence of transfer effects. In particular, it bans bidirectionality of transfer/generalisation. Suppose that two lexical items *X* and *Y* display a different syntactic behaviour in the adult-language and that such difference is due to one particular property *p*. While *X* has the value [+*p*] *Y* has the value [-*p*]. During acquisition, there may be a stage in which *X* and *Y* will behave alike if neither of them has yet acquired *p*. Consider now a situation where the child has acquired the property [+*p*] for item *X*. Provided that the relevant conditions are met, it is possible that *Y* will be assigned the value [+*p*] through Cascade-updating. At a later stage, the same item *Y* might acquire the adult setting, say [-*p*]. At this point, the Cascade-updating model rules out the possibility that [-*p*] could be transferred over to *X*. This is because, in order for [-*p*] to be added to item *X* it is necessary that the [+*p*] value be first deleted. However, as discussed in section 4.2, deletion - in the form strikethrough - is not a transferable property. This ensures that, in the scenario described above, item *X* will retain its original [+*p*] value, thus blocking bidirectional transfer. In sum, if a property is transferred at all, it will be systematically in one direction since Cascade-updating cannot override existing

information, a necessary condition for bidirectional transfer. I will return to this in chapter 2 (especially sections 2.4 and 2.5).

## 7. Concluding Remarks

In this chapter I have proposed a model of lexical updating which I claimed to consist of two mechanisms, namely Paradigm Formation and Cascade-updating. I began by discussing how the interaction between these two mechanisms leads to the spreading of newly acquired properties throughout the lexical paradigms. Moreover, I suggested that this model, though sufficiently conservative to avoid massive overshooting of the target, may allow us to accommodate overgeneralization as an inevitable consequence of the acquisition process.

I then discussed some evidence from the literature which indicates that bilinguals store lexical items from both languages in an integrated lexicon. I subsequently suggested that this may result in Paradigm Formation affecting the bilingual lexicon. With this in mind, I suggested that the mechanisms I claimed to be responsible for overgeneralization in monolinguals may also be responsible for transfer effects in bilingual acquisition. I concluded by looking at some of the implications that this hypothesis may have with regard to what types of transfer effects can be expected.

The natural next step at this point is to evaluate the empirical coverage of the model by testing some of its predictions. However, some considerations are in order. Firstly, the evaluation of any newly proposed theory can only be carried out in conjunction with background assumptions about the nature of the phenomena it investigates. This may be unproblematic in some cases. For example, whether we assume syntactic categories to be atomic or composed of sub-features (as we have done here) may be just a matter of notational preference which does not affect the ultimate outcome of the model. Other issues, however, may be highly controversial. Inevitably, evaluating a model of the type proposed here presupposes background knowledge as to what features are associated with what lexical items, as well as how these relate to the features of other elements contained in the lexicon. As is well known, this type of background knowledge is heavily dependent on theoretical conjecture (Lakatos 1973). Consequently, it may be the case that the Cascade-updating model successfully accounts for a given data set when applied in conjunction with theory  $t$  but completely fails to do so in conjunction with theory  $t1$ . This may not be a trivial problem. Of course, once a particular theory is assumed in order to account for a given set of data, then it becomes important whether the

model makes the correct predictions throughout or only in part. If the latter case applies, the model at hand is faced with immediate empirical problems. Similarly, if the model makes incorrect predictions irrespective of the hypothesis we borrow as unproblematic background knowledge, then it is probably wise to conclude that the model's design is to be altered or – possibly – abandoned altogether. On the other hand, the fact that the model makes incorrect predictions in conjunction with some specific theory as opposed to some other should not be given too much importance.

In the remainder of this thesis I will attempt to give an explanation for some more or less well-known phenomena within the field of monolingual as well as bilingual language acquisition. While doing so I will attempt to also evaluate some existing theoretical positions in order to determine which we might promote to 'unproblematic background knowledge'. However, I will inevitably have to gloss over theoretical debates – some of which are long-standing – for failing to do so would lead the acquisitionist towards a blind alley.

## NOTES

---

<sup>1</sup> This condition Andrews (1990) proposes appears to be a specific instantiation of the more general Elsewhere Principle (see Kiparsky 1973 and subsequent work).

<sup>2</sup> There are some potential exceptions to this claim. For example, some English speakers alternate between *mowed/mown* and *showed/shown* as perfect participles. Similarly, in Standard Italian there are alternations of the type *perdei/perdetti* (lose.PAST.1SG) and *visto/veduto* (see.PART.1SG), *perso/perduto* (lose.PART.1SG). Nevertheless, my impression is that although these alternations do exist within the Italian speech community, individual speakers have a strong tendency to produce only one of the two forms. Moreover, the two alternants do not seem to be in free variation even for those speakers who do use them both, register being a major player in the lexical choice. A similar case might be made for the English alternations mentioned above.

<sup>3</sup> Note that this is not equivalent to what I have just said in relation to *he walks* blocking *he walk*. In the blocking case, it is assumed that *walks* is a more specific alternative (as it's marked for person and number) and thus blocks the application of the more general *walk* (which is unmarked for both). However, there is nothing that stops the system from building a form *walk* which is also specified for Person and Number. The overwhelming absence of free variation within morphological inflections, however, suggests that this does not happen. The principle in (3b) is an attempt to formalise this.

<sup>4</sup> If we take the actual form *walk*, as opposed to some empty suffix, as being the underspecified case, then it will necessarily be the case that the same feature bundle can be associated with other elements, as virtually all other English verbs will have an unmarked form. This can be accommodated by assuming that the principle in (3a) is only active within paradigms. In doing so, we can allow, say, *walk*, *laugh*, and *smile* to be associated with exactly the same feature bundle (i.e. the 'unmarked' bundle), while still disallowing free-variation between *walk* and *walks*.

<sup>5</sup> Whether we assume the analysis in (1) to apply to the item *walk* or to a phonologically empty suffix does not affect this point.

<sup>6</sup> See Blevins (1990) for a similar argument against using negated features as an alternative.

<sup>7</sup> But see chapter 4 for discussion on how these properties can be formalised using a binary system.

<sup>8</sup> However, for ease of presentation I will occasionally make use of the terms 'lexical item' or 'lexical element' to refer to these cross-modular associations.

<sup>9</sup> The assumption that [+wh] is associated with tensed auxiliaries is not without problems. For example, it raises the question of what happens to the [+wh] feature when an auxiliary is inserted into a declarative sentence. However, my consideration of Rizzi's account is entirely for illustrative purposes.

<sup>10</sup> I include a phonemic representation in order to acknowledge the mapping onto phonological information. This is only for convenience, as I am not particularly committed to phoneme theory. Also, note that in each case I will include those properties that I consider relevant for the specific case at hand. The absence of any other property should not be taken as a statement regarding the absolute lack of such property.

<sup>11</sup> In fact, it is incompatible with any system that views language acquisition solely as the acquisition of lexical properties and their values (e.g. Chomsky 1995).

<sup>12</sup> Hale and Reiss (2003) independently pointed out the importance of this relationship in their discussion on the Subset Principle, an issue which I will address in section 2.4.4 below.

<sup>13</sup> For the purposes of this example, I will ignore semantic properties.

<sup>14</sup> By 'full' is meant that it must contain sufficient features to be recognised as belonging to a specific word category. If we adopt the binary system described above, [+V, -N] will be a full bundle while [+V] will not. The status of 'word category' as a minimal condition is hardly surprising if we consider that it is the main property the system employs when cataloguing lexical items.

<sup>15</sup> This raises the question of how many items are required before a paradigm is formed and cascade-updating is triggered. At the moment I do not have a definite answer to such question. However, I would suspect that the number need not be large at all. In fact, a generalisation such as "I said you", potentially based on the well-formedness of "I told you" seems to be drawn from the postulation of a two-word paradigm, namely *say* and *tell*. I leave the discussion for further research.

<sup>16</sup> A similar prediction with regard to gradual development would likewise follow if we took the view that acquisition proceeds on an item-by-item basis based on input frequency (see for example Maslen et al. 2004). However, according to this latter view, gradual development is a necessary outcome, since acquisition is assumed to be solely dependent on input, without the aid of any generalising mechanism. A thorough evaluation of these two competing views is postponed until further research.

<sup>17</sup> For an overview of how gradual development might be tackled within a parametric model the reader is referred to the work of Yang (2002, 2004).

<sup>18</sup> Of course, the occurrence of overgeneralisation also implies that the acquisition device is able to retreat despite the absence of negative evidence. This is a notoriously difficult issue and one on which I will have nothing to contribute.

<sup>19</sup> Although some researchers idealise away the occurrence of overgeneralisation (e.g. Fodor and Sakas 2005).

<sup>20</sup> Also, condition (ii) seems to be a rather odd property to ascribe to a biological system as it implies that language variation is a built-in characteristic rather than a surface property.

<sup>21</sup> Though the idea of 'conservative learner' informally described in some of the work on Distributed Morphology (see for example Demuth 1998) could be seen as a hint in this direction.

<sup>22</sup> This is merely for the purpose of illustrating the existence of a second subset problem. I am in no way committed to the existence of these particular features.

<sup>23</sup> This question is based on what Fodor and Sakas (2005) ask at beginning of their discussion. However the two questions differ in one crucial respect. While F&S refer to generalisation of behaviour from sentence to representation, I am looking at generalisation of lexical properties from individual item to class (or indeed subclass).

<sup>24</sup> The semantic properties ascribed to each element are meant as mnemonic devices rather than theoretical statements. See section 3.2.2 for a more theoretically motivated notation of argument structure.

<sup>25</sup> Radford (1990) points out that young children may sometimes use transitive verbs without an object. Although this might imply that children's grammar diverges from that of the adult in that it may allow optional realisation of (some) arguments,

it does not weaken my point here, for the fact remains that children do produce transitive verbs with an object, albeit not always.

<sup>26</sup> Pinker (1989) argues on learnability grounds that this assumption should be favoured unless developmental data make it untenable. However, see van Kampen (to appear) for a different view.

<sup>27</sup> Note that the mapping of one phonological string onto more than one argument/conceptual structure does not violate the Biuniqueness Principle (cf. 3 above) as this is assumed to apply only to the domain of morphological information.

<sup>28</sup> Note that the only arbitrary information that needs acquiring is the set of semantic properties, because the structure of the theta-grid follows from the universal rule in (26).

<sup>29</sup> Note that the relevant paradigmatic properties will not include theta-roles as their presence is regulated by a universal principle. See also footnote 23.

<sup>30</sup> The embedding of an event within an event is well-formed only if the embedded structure is monadic. This follows from the requirement that one of the arguments of dyadic predicates be assigned a feature in order to be formally distinguished from the other (see Reinhart 2000). What feature is assigned depends on the type of predicate and events are assigned the feature *initiator*. However, the argument of an embedding *event* is itself an *initiator*. It follows that when a dyadic *event* is embedded in an *event* the resulting structure will either (i) have more than one argument with the feature *initiator* or (ii) lack formal distinction of the two embedded arguments. Neither of these structures is well-formed. (For details of feature assignment to dyadic predicates, the reader is referred to Reinhart 2000 and Neeleman and van de Koot 2002).

<sup>31</sup> The predicate *stand* has in fact a further semantic structure which involves embedding of a state, as can be seen in the following distinction:

- (i) I stood for hours
- (ii) I quickly stood on the table



However, this is irrelevant to the present discussion and will not be considered.

<sup>32</sup> It seems obvious that *be* must have more than one conceptual structure. The one I propose here is the non-copular structure that emerges when the predicate is used in the sense of *exist*.

<sup>33</sup> The different number of arguments cannot be taken as a syntactic difference since it follows from the semantic structure. See also note 4. Also, see section 2.1 for a brief sketch of Pinker's (1984) argument in favour of constraining the postulation of phonologically empty material.

## CHAPTER 2

---

### Transfer and Delay: two sides of the same coin?

#### 1. Introduction

Towards the end of chapter 1 we looked at how the same mechanism that is claimed to be responsible for overgeneralisation in monolinguals might also explain the nature of transfer effects in bilinguals. In this chapter I will discuss some instances of transfer effects and show how they might follow once we assume the Cascade-updating model.

In sections 2 and 3 I will introduce two case-studies, both of which involve examples of non-adult utterances that have not been attested in monolingual studies. I will then show how these can be analysed as cross-language generalisations triggered by the paradigmatic similarities between the items at issue.

I will also discuss how some of the data from these two case-studies carry an important implication in relation to another property of the Cascade-updating model, namely informational monotonicity. As an informationally monotonic system, Cascade-updating is incompatible with bi-directional transfer of a given lexical property. If a property is transferred at all, it will be systematically in one direction since Cascade-updating cannot override existing information, a necessary condition for bi-directional transfer. The data I discuss in section 2.3 suggest that transfer effects are indeed unidirectional.

The data to be discussed in section 4, on the other hand, represent a case in which the bilingual child temporarily fails to develop certain elements in one of the two L1s. I will argue that although this phenomenon seems to have little in common with transfer effects, it arises from the application of the same cognitive mechanisms. I will then propose an analysis in terms of Cascade-updating and

argue that the terms 'transfer effects' and 'delay' are merely labels which refer to different surface effects of the same underlying process, namely cross-language generalisation of some property(-ies).

## 2. Transfer in WH-Elements

### 2.1 An English-Italian Case Study

#### 2.1.1 Introduction

Wh-elements in English and Italian differ with regard to their selectional requirements. Italian has two separate wh-elements which, although truth-conditionally identical, differ with regard to their ability to take a complement. While *che* must take an NP complement, *cosa* must not<sup>1</sup>:

1.     a. *Cosa studi?*  
          what study.2SG  
          'What do you study?'  
  
       a'. \**Cosa materia studi?*  
          what subject study.2SG  
          'What subject do you study?'  
  
       b. *Che materia studi?*  
          what subject study.2SG  
          'What subject do you study?'  
  
       b'. \**Che studi?*  
          what study.2SG  
          'What do you study?'

As shown in (1), these two elements behave alike in a number of respects. For example, they both refer to non discourse-given entities (i.e. they translate as *what* rather than *which*) which, following Pesetsky (1997, 2000) can be formally represented by associating them with a [ $\neg$ D-linked] setting. It is also the case that they must both move to sentence initial position in order to form a question, a property which is generally taken to result from a [+wh] value being associated with C (Chomsky 1995, see also Adger 2003).

Nevertheless, the two elements differ with regard to their selectional requirements. This can be formally expressed in terms of the setting of the feature responsible for complement selection, which I will identify here as [comp]. While *che* will be formally associated with a [+comp] specification, *cosa* will bear a [-comp] setting, thus giving rise to the following mappings:

2.

- a.     ¬D-linked           ↔     C(+wh), +comp       ↔     /ke/
- b.     ¬D-linked           ↔     C(+wh), -comp       ↔     /kɔza/

This contrasts with the English element *what* which can occur with or without a complement NP, as shown by the grammaticality of the glosses in (1). This indicates that the English wh-element does not have a specification for the [comp] feature, hence the following representation:

3.

- ¬D-linked                   ↔     C(+wh), acomp       ↔     /wɒt/

It seems that the difference between the Italian elements on the one hand and the English element on the other is encoded in the specification of one single property, namely [comp]. In cases where an English-Italian bilingual acquires the setting for one language before the other it is conceivable that Cascade-updating might apply, provided that all the necessary conditions are met. This application would then lead to the emergence of transfer effects whereby wh-elements from the two languages would behave alike. In the next section I will introduce a case-study which appears to represent this very situation.

### 2.1.2 The Subject

The data discussed in the following sections are from the case study of Lorenzo, an only child who has been exposed to English and Italian from birth. The data were collected on two separate occasions. The main collection occurred over the four-week period that goes from age 3;5 to 3;6. A second period of data collection lasted a week and took place when Lorenzo was aged 3;11.

The case of Lorenzo falls under Meisel's (1989) definition of Bilingual First Language Acquisition as he was exposed to two languages literally from day one and has been exposed to those languages with some regularity ever since. He was

born on 19<sup>th</sup> January 1998 at Meloni Hospital in Milan. His mother is a speaker of American English and, though she grew up in Kansas, her English is not particularly distinctive of any American variety as her job involved daily travel to various American States for a period of about eight years. Lorenzo's father speaks a north-western variety of Italian which can be defined as the standard variety in the Lombard region of Italy.

Lorenzo lives in Mede, a small town in the north-west of Italy where he has been addressed in English by his mother and in Italian by his father as well as almost anyone else around him. At age 1;6 Lorenzo travelled to the United States with his mother where he was exposed exclusively to English for a period of ten days.

During the first four weeks of data collection, Lorenzo and I spent an average of seven hours a day together. Data collection took place in his familiar environment. This was either in his home or at the local park, where he would be surrounded by other children of around the same age. The majority of the data was uttered spontaneously and, though some elicitation took place (see section 4.1.2), the main technique used was simple conversation. The last week of data collection involved spending less time with the child (around 3 hours a day) and this time it was in the presence of his parents. The collection method was a daily diary.

### 2.1.3 The Data

Lorenzo's speech was observed for a relatively short period of time and therefore some caution is in order when drawing conclusions from this data set. Nevertheless, while a single example may not be particularly significant, structures that were produced on two or more occasions within the four weeks may reflect a developmental pattern that deserves attention.

One such pattern that emerged from Lorenzo's data regards the Italian wh-element *cosa*. On several occasions, Lorenzo produced non-adult sentences whereby this element would be followed by a complement:

4. a. Cosa lavoro vai      a fare?      (Lorenzo 3;11;2)  
       what job    go.2SG at do.INF  
       'what job are you going to do?'
- b. Cosa cartone sai      fare?      (Lorenzo 3;11;4)  
       what cartoon know-2SG do.INF  
       'what cartoon can you imitate?'



unable to analyse *wh*-elements as capable of taking a complement. This possibility is eliminated by the fact that, most of the children<sup>2</sup>, had already begun producing *che* followed by NP complements, although sporadically:

8. Che cos' è questo? (Camilla 3;4)  
 what thing is this  
 'What is this?'
9. In che mano è? (Raffaello 2;6)  
 in what hand is  
 'In which hand is it?'

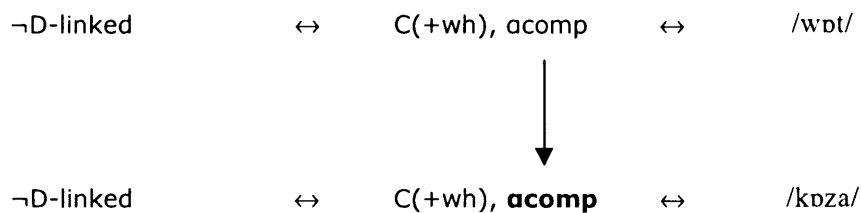
This suggests that the lack of complementation in the children's use of *cosa* was due to the presence of a negative setting for the [comp] feature, as required by the target language.

In view of these results, it is reasonable to assume that the examples in (4) are the consequence of a transfer effect whereby Lorenzo associated the English setting [a<sub>comp</sub>] with the Italian *wh*-element.

#### 2.1.4 The updating process

As the mappings in (2) and (3) show, before the acquisition of the [comp] feature the relevant *wh* elements for English and Italian share exactly the same syntactic and semantic features. Given the definition of Paradigm Formation proposed in chapter 1, the language system is expected to perceive the two items as belonging to one paradigm sharing the paradigmatic properties [¬D-linked], [+wh]. If it so happens that Lorenzo acquired the [comp] feature for the English element whilst the Italian element still lacked this feature, the right environment for the application of Cascade-updating would have arisen, thus leading to the following operation:

10.



This will then result in the Italian element behaving just like its English counterpart when inserted into the syntax. The data in (4) suggest that this is what may have happened.

#### 2.1.5 Some remarks on monolingual transfer

Because Cascade-updating is claimed to apply to the monolingual as well as the bilingual lexicon, a question arises as to whether the process in (10) might also apply to the elements in (2) since these too differ minimally. Given Paradigm Formation, if one of the two elements acquires its specification for the [comp] feature first, we might expect that this be transferred over to the other element. Such a process would then result in overgeneralisation. Indeed, analysis of the speech of two monolingual children, Guglielmo and Rosa (cf. 7 above), does suggest that some overgeneralisation had taken place.

For a period of about 6/8 months, Guglielmo and Rosa's use of *che* was not representative of the lexical specification given in (2a). Interestingly, however, this did not seem to be due to a different variety being spoken in these children's linguistic environment (cf. footnote 1). As can be seen from the following two exchanges, it seems that the child speaks a 'dialect' different from that of his parents:

#### 11. *Exchange between Guglielmo (2;2) and Mother*

- \*CHI: *che sono ?*  
          *what are.3PL*  
          *'What are they?'*
- \*MOT: *eh , dimmelo te cosa sono !*  
          *erm tell-me-it you what are.3PL*  
          *'You tell me what they are!'*

#### 12. *Exchange between Guglielmo (2;2) and Father*

- \*FAT: *lo sai       cos' è questo ?*  
          *it know.2SG what is this*  
          *'Do you know what this is?'*
- \*CHI: *eribassi cotè   bimmino ettro       butta       qui.*  
          *[unintelligible] child   [unintelligible] throw.3SG here*
- \*CHI: *e   questa che è ?*  
          *and this   what is*  
          *'And what is this?'*



13. *Exchange between Rosa (2;4) and Mother*

\*MOT: te lo sei dimenticato ?

you-CL it-CL be.2SG forgotten

'Have you forgotten it?'

\*CHI: che ha in mano que' ?

what have.3sg in hand this

'What's in this guy's hand?'

\*MOT: cos' ha in mano ?

what have.3sg in hand

'What's in his hand?'

These examples are particularly telling. In (11), Guglielmo asks a wh-question using *che* without a complement. His mother responds by repeating the same question but using *cosa* instead, thereby indirectly correcting the child. The same situation occurs in (13). In the example in (12) both Guglielmo and his father ask a zero complement wh-question. However, in spite of the fact that the father sets an example by using *cosa*, Guglielmo's wording involves using *che*. This difference in bias was reflected throughout the corpus. As the tables in (14) and (15) show, the variety spoken in both Guglielmo's and Rosa's environment has a strong bias for using *che* in contexts where a complement is required, while almost always employing *cosa* in complement-less cases.

14. *Distribution of wh-elements in Guglielmo's environment*

	<b>che<sub>prn</sub></b>	<b>che+noun</b>	<b>cosa<sub>prn</sub></b>	<b>cosa+noun</b>
Mother	3/28 (10.7%)	25/28 (89.2%)	131/131 (100%)	-
Father	1/11 (9%)	10/11 (90.9%)	63/63 (100%)	-
Sister	N/A	N/A	2/2 (100%)	-
Observer	N/A	N/A	1/1 (100%)	-

15. *Distribution of wh-elements in Rosa's environment*

	<b>che<sub>prn</sub></b>	<b>che+noun</b>	<b>cosa<sub>prn</sub></b>	<b>cosa+noun</b>
Mother	2/36 (5%)	34/35 (94.4%)	557/557 (100%)	-
Family friend	0/1	1/1 (100%)	12/12 (100%)	-
Investigator	0/3	3/3 (100%)	76/76 (100%)	-
Observer	0/6	6/6 (100%)	92/92 (100%)	-

Consequently, it seems that Guglielmo's and Rosa's use of *che* was not representative of their environment, suggesting that they had not yet achieved the target setting for this wh-element<sup>3</sup>. However, their use of *che* differed from that of the adults in a revealing manner, as it seems to have been indistinguishable from the other wh element. Compare (16)-(18) with (11)-(13) above:

16. *Exchange between Guglielmo (2;2) and Mother*

\*MOT: e questi sono bellissimi +...

and these are.3PL beautiful

\*CHI: cosa sono ?

what are.3PL

'What are they?'

\*MOT: cosa sono ?

what are.3PL

'What are they?'

17. *Exchange between Guglielmo (2;5) and Father*

\*FAT: queste sono le gambine .

these are.3PL the leg.DIM.PL

'These are the little legs'

\*FAT: hai visto ?

have.2SG seen

'Do you see?'

\*CHI: cosa sono ?

what are.3PL

'What are they?'

18. *Exchange between Rosa (2;4) and Mother*

\*MOT: *tieni.*

*keep.2sg*

*'Here you are'*

\*CHI: *cos' è ?*

*what is*

*'What is it?'*

\*MOT: *apri da sola*

*open.2sg by alone*

*'Open it by yourself'*

Since none of the above utterances involve repetition of the parent's utterance, it seems safe to conclude that they reflect the child's competence. If this is the case, it appears that what took place in Guglielmo and Rosa's lexicon was very similar to the updating process in (10). This time, however, it applied within the monolingual lexicon and it involved transferring the [-comp] specification for *cosa* over to *che*.

Having said that, it must be noted that the use of both *che* and *cosa* without a complement is also compatible with the assumption that the children had set an underspecified value for the [comp] feature. On closer inspection, however, this possibility must be excluded. If it were the case that the children had associated an [acomp] setting with both wh-elements, we expect that - during the same stage - they would also produce instances of *che* and *cosa* followed by a complement noun, since this is compatible with an underspecified [comp] value. On the other hand, if it is the case that an updating process similar to that in (10) had taken place, we only expect to find instances of complement-less wh-elements during this period. A search of the relevant files suggests that the latter case applies. Between the age of 2;0 and 2;9 Guglielmo uttered 15 instances of *che* and 25 instances of *cosa*, none of which involved a complement<sup>4</sup>. Similar results were obtained for Rosa. Out of the 38 wh-uses recorded (29 instances of *che* and 9 of *cosa*) none was followed by a complement. These results relate to the very same time-span when the utterances in (11)-(18) were recorded, thus excluding an analysis in terms of under-specification.

In sum, the acquisitional path followed by Guglielmo and Rosa provides an example of how transfer and overgeneralisation may target the very same elements. This supports the view that these phenomena are different manifestations

of the same mechanism. It also gives further support to the Cascade-updating analysis suggested in (10).

### 2.1.6 Summary

The data presented in (4) diverge from adult Italian as well as from monolingual child Italian, thus suggesting that we are dealing with instances of transfer. These data can be given a straightforward explanation by assuming the Cascade-updating model. Because the *wh*-elements in the two languages shared the same properties and no conflicting properties intervened, a feature from the English *wh*-element *what* could be transferred over to the Italian element *cosa* thus leading to cross-system generalisation. A similar analysis is also compatible with the developmental pattern exhibited by Guglielmo and Rosa, two monolingual children whose use of the *wh*-element *che* did not reflect that of their speech community.

In the next section I will discuss similar findings which have been reported in the case-study of a child acquiring a language pair different from the one just presented.

## 2.2 Cantonese v. English

### 2.2.1 Typology

As is well known, Cantonese *wh*-elements do not undergo syntactic movement. This is shown in (19) (from Yip and Matthews 2000:195)<sup>5</sup>:

19.    Lei sik-zo    matje?  
        you eat-PERF what  
        'what did you eat?'

This contrasts with English *wh*-elements which are typically analysed as moving out of the object position and into the C domain:

20.    [CP What<sub>i</sub> [c did<sub>k</sub>][IP you [I t<sub>k</sub>][VP eat t<sub>i</sub>]]]

As already mentioned in section 2.1.1, this configuration is generally taken to result from feature strength associated with a property of C (Chomsky 1995, Adger 2003). By hypothesis, in a language like English the [*wh*] feature bears a plus value. In contrast, the Cantonese *wh*-elements are associated with a [-*wh*] feature and therefore do not need to – and by Economy cannot – raise overtly (Chomsky 1995). Translated into mapping principles, this yields the following representations:

21.

a.  $\neg$ D-linked  $\leftrightarrow$  C(-wh), acomp  $\leftrightarrow$  /matje/b.  $\neg$ D-linked  $\leftrightarrow$  C(+wh), acomp  $\leftrightarrow$  /wɔt/

Note that, as far as complementation is concerned, the Cantonese wh-element behaves identically to the English element (but unlike the Italian one, cf. 1), as shown below (from Yip and Matthews 2000):

22. a. Lildou matje aa?

here what PART

'what is in here?'

b. Lil go matje ngaansik aa?

this CLS what colour PART

'what colour is this?'

Consequently, the setting for the Cantonese [comp] feature in (21a) above is given as underspecified just like in the English counterpart.

### 2.2.2 Paradigm Formation

Although the mappings given in (21) may not be exhaustive as far as the adult language is concerned, it is plausible that at the early stages of acquisition the only information available to the child may be with regard to the most prominent features, namely complementation, wh-force and D-linking. If this is correct, the specifications in (21a) and (21b) provide a potential representation of the mappings in the mind of Cantonese and English children respectively. This has an important consequence with regard to Cantonese-English bilinguals.

Given the formulation of Paradigm Formation, if the two elements co-exist as lexical mappings (as in the case of a Cantonese-English bilingual) prior to the acquisition of the wh-feature, they will both be perceived as belonging to the following paradigm:

23.  $\neg$ D-linked, C( ), acomp

### 2.3.1 Introduction

### 2.3.2 The Data

24. a. This one what? (2;04)  
b. You go to the what? (2;05)  
[sitting in the car asking Daddy]  
c. This what colour? (2;10)

25. A: I am going to the demonstration  
B: You are going to the what?

79

CHILDES database (MacWhinney and Snow 1990, MacWhinney 2000). The files analysed are as follows:

26.

	<b>Age range</b>	<b>Files</b>
Eve	1;6 – 2;2	eve01.cha – eve18.cha
Sarah	2;3 – 3;8	sarah01.cha – sarah074.cha

The overwhelming majority of in situ wh-elements found in these corpora are repetitions of adult utterances. Two examples are given below:

27. MOT: he's doing what ?  
 EVE: he doing what ? (1;08)

28. FAT: we did a what ?  
 SAR: yyy # did a what ? (2;11)

Other examples involved adult-like structures such as PPs with a wh object (29) and elliptical interrogatives (30):

29. SAR: uh nuns ? (3;4)  
 MOT: yeah .  
 SAR: for what ?  
 SAR: for what ?  
 MOT: to make money

30. SAR: you know what ? (3;5)  
 SAR: I has a tele(phone) +..

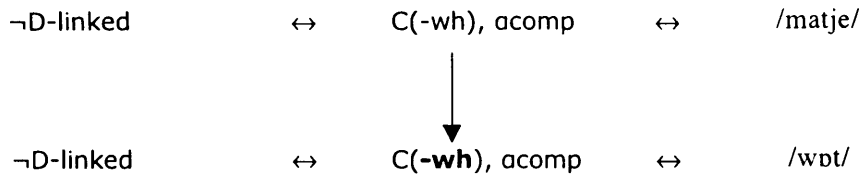
These contexts are radically different from those of Timmy's utterances. As for spontaneous echo questions, none occurred in Eve's speech at all, while Sarah did produce some but these only began to emerge around age 3;8. This seems to indicate that such structures may develop rather late, thus casting doubt on the idea that Timmy's utterances were a developmental effect due to the presence of echo questions in the English input.

Another piece of evidence that points in this direction comes from a quantitative analysis. Peng (1998, cited in Yip and Matthews 2000) compared

Timmy's rate of *in situ* wh-elements with that of Eve and found a striking difference. While only 1.6% of Eve's questions had an *in situ* element, Timmy's rate was as high as 67.6%.

In sum, it seems reasonable to conclude that the utterances in (24) are genuine instances of transfer. As discussed above, this can be given a straightforward explanation by assuming the Cascade-updating model. Because the wh-elements in the two languages share a number of properties, Paradigm Formation will perceive them as belonging to the same paradigm (cf. 23). As soon as the Cantonese element is associated with the [-wh] feature, Cascade-updating will be triggered resulting in the following operation:

31.



Once the [-wh] feature is associated with the English element, the bilingual child will produce *in situ* wh-elements in English as well as Cantonese. The application of Cascade-updating, with consequent triggering of the operation in (31), provides us with an explanation for the high number of *in situ* wh-elements in Timmy's English.

However, the operation in (31) implies that Timmy should be unable to produce sentences involving moved wh-elements. As transpires from the discussion above, this is not so, as 32.4% of the wh structures in Timmy's English involved wh-movement, and are thus incompatible with a [-wh] setting. While this serves as yet another reminder that parameter setting is not instantaneous, and thus may not lead to across-the-board effects, following a discussion in chapter 1 section 2.4.3, I have chosen to idealise away from the issue of gradual development and I will therefore not address it here.

### 2.3.3 Remarks on Unidirectionality

By definition, Cascade-updating involves the attribution of a property to an item which lacks a value for that property. Moreover, Cascade-updating is an informationally monotonic mechanism and thus unable to overwrite any information. This means that – for a given property – if lexical items do affect each other, they are expected to do so systematically in one direction. This leads to the prediction that Timmy's Cantonese should lack instances of moved wh-elements.



Indeed, Yip and Matthews (2000) report that no examples of wh-moved elements were found in Timmy's Cantonese data. This lack of bi-directional transfer gives further support to the Cascade-updating analysis proposed here.

However, Kwan-Terry (1986) reports data from another Cantonese-English case-study where she suggests that bi-directional transfer may have occurred. I will discuss this below.

## **2.4 A Second Cantonese-English Case Study**

### *2.4.1 Introduction*

Kwan-Terry (1986) discusses the case-study of a Cantonese child, Elvoo, who was also exposed to English from an early age. The corpus includes data from between the age of 3;6 and 5;0. The collection method was audio-recording plus observational data.

It must be pointed out that, unlike the cases discussed so far, Elvoo's case does not fall under the definition of BFLA as he was not exposed to both languages from birth. Up to the age of 1;0, he was exposed only to Cantonese, this being the language of his parents as well as the community. Moreover, although English was introduced into the household when Elvoo was 1 year old, this did not involve exposure to native speakers. The reason why English came to be spoken in Elvoo's home was due to the arrival of a non-Cantonese speaking Filipino maid who, as Kwan-Terry points out, had "just enough simple English to enable her to communicate with other members in the household" (1986:15). The other speaker who provided English input was Elvoo's sister, Poupee. Six years Elvoo's elder, Poupee had been exposed to native English at the age of 5;0 when she stayed in England for about one year. Although she continued to improve her English after returning to Singapore, this involved a more formal training rather than natural exposure.

Given these facts, it is not at all obvious that the Cascade-updating model should be able to account for Elvoo's case. Although the model may have the potential to be expanded so as to explain some features of second language acquisition or even second language learning, crucial differences between these areas of study need to be carefully considered before any claim can be made as to the feasibility of such an undertaking. Nevertheless, Elvoo's case is of interest to us in so far as it provides grounds for discussion in relation to the claim made in section 2.3.3, namely that bi-directional transfer is predicted to be absent cross-linguistically. As in the testing of any theoretical prediction, it is important to consider what kind of data would provide counterexamples. In this sense, Elvoo's

case is important as I believe it does not straightforwardly provide examples of bi-directional transfer, despite claims to the contrary. As I will argue below, the examples that have been analysed as non-adult Cantonese sentences may just be instances of regular development which have nothing to do with interference from English.

#### 2.4.2 *The Data*

Between the age of 3;6 and 3;9, Elvoo went through a stage during which he produced English utterances containing *in-situ* wh-elements (from Kwan-Terry 1986:23):

- |     |                             |       |
|-----|-----------------------------|-------|
| 32. | a. You are doing what?      | (3;6) |
|     | b. You are eating what?     | (3;9) |
|     | c. This is for making what? | (3;9) |
|     | d. Your name is what?       | (3;9) |

Arguably, the majority of Elvoo's interrogative utterances involved *in-situ* wh-elements, the only target-like utterances he produced being potentially analysable as unstructured chunks. These systematically involved a contracted form of the copula which was almost always followed by a demonstrative (from Kwan-Terry 1986:22):

- |     |                 |       |
|-----|-----------------|-------|
| 33. | a. Who's that?  | (3;8) |
|     | b. What's that? | (3;6) |
|     | c. What's this? | (3;7) |

Although Kwan-Terry (1986) does not provide percentages, it appears that the vast majority of Elvoo's novel interrogative utterances involved leaving the wh-element *in-situ*.

At around 4 years of age, pre-posed wh-elements began to emerge in Elvoo's speech (from Kwan-Terry 1986:24):

- |     |                        |       |
|-----|------------------------|-------|
| 34. | a. Where I will hide?  | (4;3) |
|     | b. Where you got this? | (4;3) |

So far there seems to be nothing unsurprising about Elvoo's development. The data in (33) indicate that he went through the same stage as Timmy (cf. 24), while (34) simply shows that he eventually retreated and began to move towards the target.

However, Kwan-Terry (1986:25) reports that – a few months later – “his Cantonese wh-questions became affected”. This conclusion is based on the presence of the following utterances in Elvoo's speech:

35. a. *Matje lei zungji* (4;9)  
      what you like  
      b. *Matje lei ju maai* (4;9)  
          what you want buy  
      c. *Bingo lei zungji* (4;10)  
          which you like

Although these utterances diverge from adult Cantonese, the conclusion that they result from transfer is – as always – bound to the lack of such instances in the speech of monolingual children. Not until we have established their absence in monolingual development can we conclude that they are the result of transfer. Nevertheless, Kwan-Terry (1986) does not offer any comparison with monolingual Cantonese children. This becomes of even greater importance when we consider that adult Cantonese does allow pre-posing of some wh-elements, notably *matje* ('what') and *bingo* ('who'), in contexts where they are topicalised. This immediately raises the question of whether Elvoo was simply in the process of acquiring this topicalisation rule. A further reason for pursuing this line of explanation is given by Kwan-Terry herself. Topicalisation of wh-elements, she reports, typically does not involve *bin* ('where'). Interestingly, no examples were found in Elvoo's speech which involved pre-posing of *bin* ('where'). However, without more information about the monolingual development of this topicalisation structure it is impossible to draw conclusions in either direction. In the next section I will discuss some research which casts further doubts on the idea that the utterances in (35) are instances of transfer.

## 2.5 Monolingual Cantonese

### 2.5.1 Introduction

Wong et al (2004) carried out an experiment to assess the development of wh-questions in children acquiring Cantonese as their first language. Twenty children participated in this study<sup>6</sup> and they ranged in age between 3;4 and 5;8. The



The experiment involved a total of ten trials per child plus two control trials to ensure that the child was paying attention.

### 2.5.3 Results

The older children produced grammatical wh-object questions 77.4% of the time. For the younger children, however, the results were striking: only 35.2% of their questions were target-like. Even more surprisingly, the rate of responses that involved pre-posing of the wh-element was as high as 63%. Some examples are given in (38) below (from Wong (p.c.) and Wong et al 2004:1449-1450).

38.	Target	Child's response <sup>8</sup>
a.	hungzai pou hei bingo aa bear pick up who PRT	bingo pou hei hungzai aa who pick up bear PRT
b.	hungzai laai bingo bear pull who	bingo tung hungzai laai jan who with bear pull person

It was reported that the structure in (38a) accounted for 39% of the responses, while 24% of the responses involved the structure in (38b). The remaining errors did not involve wh placement<sup>9</sup>.

### 2.5.4 Analysis

The experimental results indicate that monolingual Cantonese children go through a stage when they consistently produce interrogative sentences with fronted wh-elements<sup>10</sup>. This casts serious doubts on the claim that Elvoo's errors were instances of bi-directional transfer (cf. 35).

A type of transfer that does seem to have occurred in Elvoo's case, however, is in relation to the verbal domain. Besides the occurrence of wh-movement, the monolingual errors in (38) also involve movement of the main verb to the C domain<sup>11,12</sup>, a feature which was not reported for Elvoo's speech. This is clear when comparing (38a) with (35). Although the reasons behind verb movement in monolingual Cantonese children are unclear, what seems to have happened in Elvoo's case is that he acquired the feature responsible for leaving the main verb *in situ* as a consequence of transfer from English. This view finds support in the data in (39) which show that – at the time when he produced the sentences in (35) – Elvoo also produced English sentences which correctly lacked movement of the main verb to C:

39. a. Where you got this? (4;3)  
 b. What you want to do? (4;6)  
 c. So what we'll make then? (4;9)

This can be explained if we assume that Elvoo's Cantonese lacked a value for the [C] feature, the feature traditionally taken to be responsible for this type of movement. As soon as he acquired the [-C] value for the English verbs, this would have been transferred over to some members of the Cantonese verbal paradigm, in the following fashion:

40.



As a result, not only would Elvoo's Cantonese involve the pre-posing of wh-elements – a property that is typical of children acquiring Cantonese (cf. section 2.5.3) – but it would also involve lack of verb movement to C.

In conclusion, Elvoo's case cannot be used as evidence in favour of bi-directional transfer. As we have seen, Cantonese monolinguals also go through a stage which involves movement of wh-elements, thus indicating that Elvoo's speech was simply an instance of regular linguistic development. However, when comparing Elvoo's utterances with those of his monolingual peers it becomes apparent that they differ with regard to verb placement. On the one hand, monolingual Cantonese children were found to produce interrogative structures that involve movement of the main verb to the C domain. On the other hand, Elvoo's speech showed no sign of such operation. This discrepancy can be explained by assuming that Elvoo had transferred the [-C] feature from English to Cantonese, thus leading to the production of interrogative structures that lack verb movement. Moreover, the data in (32) suggest that – like Timmy – Elvoo had also transferred the [-wh] feature from Cantonese over to English. Crucially, however, these operations do not involve bi-directional transfer.

The experiment reported in Wong et al (2004) provides important evidence in relation to the development of monolingual Cantonese children and it once again brings out the importance of monolingual comparison in assessing the potential

occurrence of transfer effects. In view of these findings, the assumption that Elvoo's errors involved bi-directional transfer is considerably weakened. Moreover, research needs to be done in relation to the acquisition of topicalised wh-elements in Cantonese in order to establish whether this may have a part to play in the explanation of Elvoo's developmental pattern.

## **2.6 Summary**

In this section we discussed some non-adult structures reported in three separate bilingual case-studies. Based on comparison with monolingual children from the CHILDES database, it was suggested that these structures are instances of transfer and that they can be explained by assuming the Cascade-updating model. It was then argued that very similar effects have also been found in the speech of monolingual children, a welcome result for a model that takes transfer and overgeneralisation as underlyingly equivalent (this issue will be explored again in chapter 3).

Finally, it was argued that no evidence of bi-directional transfer was found in the two English-Cantonese case studies. Data from Kwan-Terry (1986) was reanalysed in view of recent experimental work with monolingual Cantonese children and it was concluded that these represent two separate transfer effects rather than bi-directional transfer of the same property, as previously claimed.

Before moving on, I would like to clarify one last point with regard to the idea of unidirectionality. It is important to note that – within the Cascade-updating model – unidirectionality of transfer need not be dependent on the language pair. In cases where the two target languages have a conflicting value for a given property (i.e. [+P] v. [-P]) the direction of transfer will depend on which value is acquired first. While Yip and Matthews' child displays transfer effects from the Cantonese to the English paradigm, another child acquiring the same language pair could well display exactly the same transfer in the opposite direction (*modulo* restrictions on the feature at issue cf. section 2.5.4). In other words, directionality of transfer need not be constant across bilinguals. More research on English-Chinese bilinguals is needed in order to test this possibility.

### 3. Negation and Modality

#### 3.1 Transfer of Modality

##### 3.1.1 Introduction

In this section I will look at some errors involving the positioning of negated modals in the speech of Lorenzo, the English-Italian bilingual introduced in section 2.1.2. I will argue that these errors must be analysed in terms of transfer effects since they are absent in the speech of monolingual Italian children. I will then suggest how this may be done by assuming the Cascade-Update model in conjunction with some traditional analyses of modality.

##### 3.1.2 Background

It is well known that negation of modal verbs has at least two separate interpretations, namely *epistemic* and *deontic* (but see Palmer 1990 for extensive discussion of other interpretations). For a modal verb like *can*, however, a third interpretation is also available, namely *ability*. In English, this can express the presence of a skill or, if negated, the absence of it:

41. The Spice Girls can't sing

This contrasts with languages like Italian where only a deontic/epistemic distinction can be made. The following examples illustrate this difference:

42.

- |   |           |                  |
|---|-----------|------------------|
| a. I can drive (I learnt how to)          | ability   | English/*Italian |
| b. I can drive (I have a driving license) | deontic   | English/Italian  |
| c. I can drive (I consider it an option)  | epistemic | English/Italian  |

In order to express the concept in (42a), Italian makes use of alternative strategies among which are the verbs *sapere* ('know') and *riuscire* ('manage') as well as the adjectival construction *essere capace di* ('to be able to'):

- 43.
- |              |                   |                                    |
|--------------|-------------------|------------------------------------|
| a. Non so'   | guidare           | (I never learnt)                   |
| not know.1SG | drive.INF         |                                    |
| b. Non sono  | capace di guidare | (I never learnt)                   |
| not be.1SG   | able of drive.INF |                                    |
| c. Non posso | guidare           | (I've been banned/*I never learnt) |
| not can.1SG  | drive.INF         |                                    |



Like the English counterpart, *potere* may also have an epistemic interpretation, but this is ruled out in (43c) by the position in which the negative marker surfaces. Although English has been argued to employ different underlying structures for negated deontic and negated epistemic (see for example Cormack and Smith 2000), this has no effect on the linear order of the [modal neg] sequence. This clearly differs from what we find in Italian<sup>13</sup> (underlying indicates stress):

44.    a. I can't/cannot go            (=not possible)  
       b. I can not go            (=possible not)
45.    a. Non posso andare            (=not possible)  
             not can.1SG go        .INF  
       b. Posso non andare        (=possible not)  
             can.1SG not go.INF

Consequently, the example in (43c) does not allow the epistemic interpretation since Italian modals cannot take scope over a preceding negative marker.

Other interpretations may be available but these all seem to follow from a deontic reading for – in some sense – they communicate general permission. To paraphrase Lyons (1977), the deontic use of *potere* is concerned with the possibility of an act as performed by morally responsible agents. This can lead to a number of implications. For example, it could be that the speaker has a prior arrangement or that s/he has a broken leg and therefore it would be inappropriate or unreasonable for them to drive. Crucially, none of these interpretations question the speaker's ability to potentially carry out the driving action, though the circumstances might be such that their ability is hindered. In other words, the speaker is saying that there are reasons why s/he can't do what s/he knows how to do.

This highlights an important difference between the ability and the deontic reading. While knowledge – loosely speaking – is a matter of presence versus absence, circumstances are much more subjective. Whether having broken one's leg entails not driving is largely a matter of opinion (how much does the plaster hinder the driving/driving safely?) or attitude (the context in (43c) could have involved uttering "I am not supposed to drive" had the speaker's attitude towards the ban been different). Arguably, this is what modality is; the expression of an opinion/attitude as opposed to a property. For this reason, some authors (see for example Lyons 1968; Steele 1975) have pointed out that – in its ability reading –

*can* is not a modal verb at all as it simply refers to a characteristic of the subject of the sentence, just like most verbs do<sup>14</sup>. If this is the case, the difference between English and Italian has a straightforward explanation: while English *can* has both a modal and a lexical component, Italian *potere* is only modal.

### 3.1.3 Data Analysis

Turning to our case study, it is interesting to note that, on a number of occasions, Lorenzo uttered sentences which involved non-adult use of the Italian modal verb *potere* with an ability interpretation:

46. a. Non posso andare avanti  
not can.1SG go.INF forward  
(uttered while struggling with pushing his bike forward)
- b. Non puoi vedere?  
not can.2SG see.INF  
(asking whether the observer was able to see over a fence)
- c. Non posso riuscire  
not can.1SG manage.INF

The sentences in (46a) and (46b) are ungrammatical in the intended interpretation while (46c) is a non-target utterance altogether, as *potere* cannot serve in support of *riuscire*<sup>15</sup>.

It must be noted, however, that in some cases the use of *potere* with a capability meaning is allowed in Standard Italian, as in (47).

47. Non può farcela  
not can.3SG make.it  
'He can't make it'

However, this is very rarely, if at all, encountered in the Italian spoken in Lombardy, the area where Lorenzo's hometown is located<sup>16</sup>. Moreover, a search of the Italian corpora available on the CHILDES database revealed no uses of this type in the speech of monolingual Italian children of comparable age. This strongly suggests that the utterances in (46) cannot be attributed to a regular developmental stage, and must therefore be otherwise explained.

The monolingual files were taken from the Antelmi and the Calambrone corpus (Antelmi 1992, Cipriani et al 1989). The analysis included those children

whose age was closest to that of Lorenzo's at the time of recording. Relevant files are summarised below:

## 48. Italian Monolingual Files

Child	Age range	Files
Camilla	2;6 – 3;4	cam206.cha – cam304.cha
Guglielmo	2;7 – 2;11	gug5.cha – gug9.cha
Rosa	2;7 – 3;3	rosa05.cha – rosa21.cha

As far semantic properties are concerned, all the occurrences of *potere* found in these data reflected the uses reported in section 3.2.1 above. The majority of examples involved a deontic use with a request/permission interpretation:

## 49. Camilla (2;2)

- \*CHI: posso pigiarli ?  
 can.1SG press.them-CL  
 'Can I press them'  
 \*DON: no, non puoi toccare niente...  
 no not can.2SG touch nothing  
 'No, you can't touch anything'

## 50. Elisa (1;11)

- \*ELI: posso accende(r)e ?  
 can.1SG switch  
 'Can I switch it on?'  
 \*MOT: sì # e adesso lo accendiamo # senti.  
 yes, and now it switch.1PL listen.2SG  
 'Yes, now we'll switch it on, listen..'

## 51. Marco (2;3)

- \*MAR: po(sso) [/] posso gonfiare a fiato ?  
 can.1SG can.1SG blow at breath  
 'Can I blow them up by mouth?'  
 \*MOT: no non puoi gonfiare a fiato, son(o) duri [/] duri questi  
 no not can.2SG blow at breath are.3PL hard hard these  
 palloncini.  
 balloons-DIM  
 'You can't blow up these balloons by mouth, they are too tough'

The examples in (49) and (50) represent the use most commonly found in the corpus, namely permission. As to the example in (51), some remarks are in order. At first, this might seem like an ability use since the child may appear to be referring to his ability to blow up the balloons by mouth. However, when we look at the preceding context, it suggests that this is not the case. Details aside, the discussion developed as follows:

52. *Exchange between Marco (2;3) and mother*

- \*MOT: cosa stai facendo Marco?  
 what stay.2SG doing Marco  
 'What are you doing Marco?'
- \*MAR: gonfian(do) i palloncini.  
 blowing the balloons.DIM.PL
- \*MOT: non riesci son(o) duri Marco da gonfiare così a bocca.  
 not manage.2SG are.3PL hard Marco of blow.INF so at mouth  
 'But you can't, they are too hard to blow up by mouth'
- \*MAR: come gonfiamo?  
 how blow.1PL
- \*MOT: ci vuole la pompa ma sai che non so dove sia  
 CL want.3SG the pump but know.2SG that not know.1SG where is  
 andata  
 gone  
 'You need a pump, but I don't know where it is'
- \*MOT: eh@i a fiato é un po' dura.  
 INT at breath is a bit hard  
 'It's a bit too hard by mouth'
- \*MAR: un po' duro.  
 a bit hard
- \*MOT: non riesce neanche la mamma a fiato.  
 not manage.3SG not-even the mum at breath  
 'Not even mum can do it by mouth'

At first, the mother sees Marco struggling with trying to blow up the balloons by mouth and she comments that he is unable to do so because they are too hard. In this instance she is clearly referring to the child's inability, hence the verb *riuscire* (see footnote 15). She then makes a general comment, marked by the use of the impersonal construction *ci vuole* "one needs", with regard to the need for a pump.

This marks a switch of topic from Marco's inability to general impossibility whereby she points out that it is impossible for these particular balloons to be blown up by mouth. This is further clarified in the last line where she maintains that even she can't do it, thus pointing to a general (im)possibility. At this point, after a brief discussion about where the balloons had been bought, Marco utters (51). Although it may be unclear whether he was asking for clarification (hence a possibility reading) or permission, he did not evoke ability. This differs sharply from the examples uttered by Lorenzo. In (46a), Lorenzo was referring specifically to a personal characteristic of his (i.e. lack of sufficient strength) while in (46b) he refers to a characteristic of the observer's (i.e. insufficient height). These are not compatible with a general possibility reading.

In short, the most common use of *potere* in the monolingual corpora was with a permission reading. Though there were some utterances that might have involved possibility, these were far less common. This echoes many a study in the literature which found that the development of epistemic modality follows that of deontic modality (see for example Wells 1979 for English, Stephany 1986 for a crosslinguistic study which included Italian).

### 3.1.4 A Case for Transfer

Given the monolingual data, it is reasonable to conclude that the examples in (46) are a consequence of lexical transfer which – following the discussion in section 3.1.2 – appears to have involved associating the non-modal component *ability* with the Italian modal *potere* by generalising the lexical specifications of English *can*.

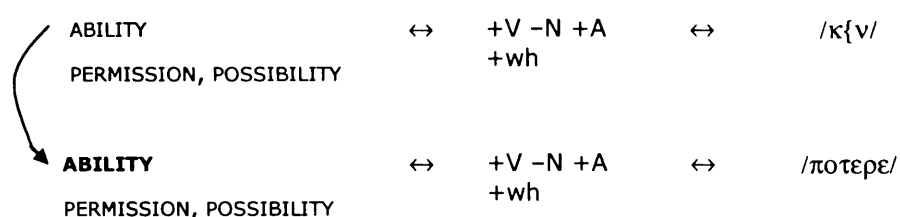
In order to see if this is compatible with the Cascade-updating model we must look at the lexical properties of these two items to establish whether they can be perceived as part of the same paradigm. This seems to be the case. Both items are auxiliaries, and they also both bear a syntactic property that requires them to move above IP/TP in interrogative contexts (Belletti 1990, Roberts 1993). Consequently, the formal difference between *can* and *potere* involves only one property, namely [ABILITY]:

53.

a.	ABILITY PERMISSION, POSSIBILITY	↔	+V -N +A +wh,	↔	/kæn/
b.	PERMISSION, POSSIBILITY	↔	+V -N +A +wh	↔	/potere/

As a result, prior to the acquisition of [ABILITY] the two lexical elements will be perceived as belonging to a single paradigm based on their hosting of the features [+V, -N, +A] [+wh] [PERMISSION, POSSIBILITY]. When the property [ABILITY] comes to be associated with the English element, Cascade-updating is triggered, giving rise to the following effect:

54.



The utterances in (46) would then be expected. Note that, if this analysis is correct, it once again raises the question of whether it is possible for transfer/overgeneralisation to be triggered by a single pair of items (see also chapter 1 footnote 15).

### 3.1.5 Modal Properties

Although *can* allows for three separate interpretations, namely ability, permission, and possibility, it is not obvious that these should follow from three separate semantic properties in the way that the specifications in (53) suggest.

On the one hand, the fact that the ability interpretation is only available in some languages, while other languages associate it with a separate lexical verb, seems to imply that ability is indeed tied to a particular semantic specification (cf. section 3.1.2). On the other hand, a representation that specifies permission and possibility separately implies that these two specifications doubly dissociate and therefore that it should be common for languages to associate each of them with different mappings. Nevertheless, deontic and epistemic modality seem to be intimately linked cross-linguistically. As Perkins put it: "a single form, or set of forms [...] is used to express more than one modal meaning not only in English, but also in Basque, Classical Aztec, French, German, Italian, Kapampangan, Korean, Luiseño, Polish, Tamil, Thai, Tzeltal, Welsh and many ancient Indo-European languages." (Perkins, 1982:246).

It is perhaps due to this crosslinguistic fact that much work has been dedicated to the idea that each modal has only one fundamental meaning (see Joos 1964, Bouma 1975, Perkins 1982, among many others). For example, Sweetser (1990) suggests that the core meaning of *can* may be paraphrased as "there is no barrier which prevents x from happening". This barrier can then be understood as a 'physical/mental' barrier, hence the interpretation in (42a), or as an 'authority', hence the interpretation in (42b). The epistemic interpretation in (42c) arises when the barrier is interpreted as being imposed by a certain 'world structure', i.e. the world is structured in such a way that no barrier prevents x from happening (see also Ehrman 1966 for a similar view). On this view, the crosslinguistic facts with regard to (42b) and (42c) are easily accounted for since the different interpretations are implicatures arising from a single basic meaning. What is problematic for this type of analysis is the ability interpretation.

If all three interpretations in (42) are derivable from the root meaning, then the ability reading should be available cross-linguistically since it is supposedly arrived at via implicature. Nevertheless, we have seen that (42a) is not a legitimate interpretation in Italian. This is also true in a number of other languages (see Palmer (2001) for a detailed discussion).

The same problem arises with semantic approaches that take modal interpretations to stem from the application of a semantic function that derives 'possible world' interpretations when fed with a particular discourse context (see for example Kratzer 1977). If a modal function can be fed three different contexts in English (hence interpretations 42a, 42b, and 42c), it is not obvious why it should only accept two of the three contexts in Italian.

A more promising alternative has been discussed by Klinge (1993). Like Joos (1964) and Bouma (1975), she proposes that the different interpretations of *can* are arrived at via enrichment. However, I would like to suggest that her system allows for a more fruitful distinction.

She argues that modal verbs are operators which encode procedural information that specify "how the conceptual information carried by the [proposition] is to be processed" (1993:321). In the case of *can*, this information has the semantic property of POTENTIALITY, which takes scope over a proposition<sup>17</sup>. Consequently, a sentence like 'John can swim' has the following representation:

55. POTENTIAL [John swim]

The idea is that the representation in (55) yields an interpretation whereby 'John' has the POTENTIAL to be the AGENT of the EVENT 'swim'. As a consequence of this under-determinacy, the following two assumptions become available to the hearer:

56.   a. There is an *internal* motivation for why John has such potential  
      b. There is a *non-internal* motivation for why John has such potential

The three meanings traditionally associated with *can* - namely ability, permission and possibility - can then be derived via a process of enrichment<sup>18</sup>. The assumption in (56a) yields the interpretation synonymous with 'John is able to swim' while the assumption in (56b) yields one of two interpretations, namely 'John is allowed to swim' and 'it is a possibility that John swims'. Which of these two interpretations is favoured depends on the context, though they are both compatible with a non-internal motivation, the only difference being whether the external entity be some sort of 'authority' as opposed to 'world structure'.

This approach offers a rather elegant explanation whereby the various meanings of *can* follow from its poor semantic specification. However, it still faces the problem of predicting that – contrary to fact – the interpretations in (42) should be crosslinguistically associated with a single form.

Unlike the proposals we have seen so far, however, this particular view might be able to overcome the problem if we allow for some adjustments. Suppose that the property POTENTIALITY is subject to specification and that such specification involves setting a feature, call it [internal], as either + or -. The crosslinguistic differences outlined in (42) can now be made to fall out. English *can* allows three interpretations because it is underspecified for POTENTIALITY while the Italian modal *potere* is specified as *-internal* and therefore it only allows the interpretations in (42b) and (42c), the desired result.

An important consequence of this proposal is that the deontic/epistemic interpretations are related to each other in a way that the ability interpretation is not. Because both the deontic and the epistemic interpretation reside in the *non-internal* motivation, they are essentially two sides of the same coin: if one interpretation is available, then the other may follow<sup>19</sup>. This seems to be a welcome result given the cross-linguistic connection between deontic and epistemic modality. The 'ability' interpretation, on the other hand, is incompatible with a modal whose POTENTIALITY is restricted to *non-internal*, thus leading to the cross-linguistic differences outlined in (42).



This move has one further benefit; the overlap between English *can* and *be able to* can be captured by assuming that *be able to* is the mirror image of Italian *potere*, namely it is specified as *+internal* POTENTIALITY, hence the following contrast:

57. a. John is able to swim. i.e. he has learnt how to (internal motivation)  
 b. \*John is able to swim. i.e. he's allowed (non-internal motivation)

In sum, expanding on Klinge's (1993) proposal, the difference between *potere* and *can* can be put down to semantic specification: while the former is specified for internal POTENTIALITY, the latter is not. This difference will be reflected in the respective mappings, as below:

58.

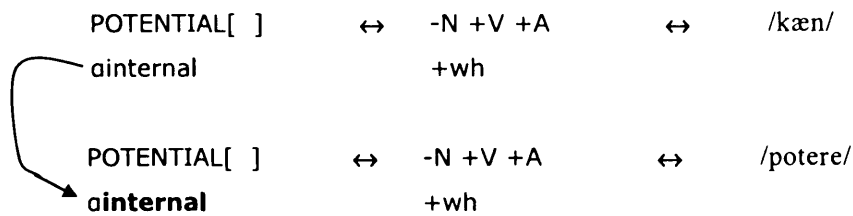
- |    |              |   |          |   |          |
|----|--------------|---|----------|---|----------|
| a. | POTENTIAL[ ] | ↔ | -N +V +A | ↔ | /potere/ |
|    | -internal    |   | +wh      |   |          |
| b. | POTENTIAL[ ] | ↔ | -N +V +A | ↔ | /kæn/    |
|    | αinternal    |   | +wh      |   |          |

At this point, the workings of the transfer effect underlying the utterances in (46) seem rather straightforward. Prior to the acquisition of the [internal] feature, the two mappings in (58) necessarily form a paradigm, having the following paradigmatic properties:

59. POTENTIAL[ ] ↔ -N +V +A

As a result, whichever mapping happens to acquire the feature [internal] first will affect the other. In Lorenzo's case, possibly due to the fact that his English was slightly ahead of his Italian, it so happened that the English item acquired the [α internal] specification first, thus leading to the following transfer effect:

60.



This erroneous association of the English specification with the Italian modal resulted in the appearance of the utterances in (46) in Lorenzo's speech.

### 3.1.5 Summary

In this section it was suggested that we can provide a plausible explanation for the use of modals in the case study presented in section 2.1.2 by assuming the Cascade-updating model in conjunction with some existing theories of modality.

Some problems related to traditional analyses of modality were also briefly discussed and it was suggested that a revised version of Klinge's (1993) analysis may be the preferred alternative. Finally, it was shown how the Cascade-updating model, in conjunction with Klinge's analysis, could help us understand the transfer effects in (46).

In the next section I will discuss further data from the same case-study and consider whether these, too, are compatible with the Cascade-updating model.

## 4. Transfer as Delay

### 4.1 Introduction

While in sections 2 and 3 above I discussed how the application of Cascade-updating can result in the appearance of transfer effects, in this section I will present the possibility that cross-language generalisation might have a different outcome. More precisely, I will look at how transfer of a property may result in one of the two languages temporarily failing to develop certain elements. When this occurs, the bilingual child seems to lag behind when compared to monolingual children of the same age.

Within the traditional literature, this phenomenon has been considered to be unrelated to transfer effects. For example, Paradis and Genesee (1996) suggest that delay might occur due to the putative difficulty that bilingual children have in dealing with two languages. Although Paradis and Genesee (1996) do not offer a

detailed explanation for this claim, they seem to imply that the bilingual child might put the development of some elements on hold due to his/her linguistic system being overloaded.

In this section I argue that the underlying causes of delay are in no way different from the causes of transfer effects as discussed in sections 2 and 3 of this chapter. Like transfer effects, delay is caused by the transfer of a lexical property over to the other L1. However, it might so happen that the resulting element cannot be integrated into the grammar for independent reasons, such as lack of a certain supporting structure. Importantly, however, no extra assumptions need to be made about the computational abilities of bilingual children. Consequently, I will take the position that, although a potentially useful shorthand, the term delay does not refer to any particular underlying process. Far from being due to “extra burden” (Paradis and Genesee 1996) that is put on the shoulder of the bilingual, delay is a consequence of cross-language generalisation, as conceived of within the Cascade-updating model<sup>20</sup>.

#### 4.1.1 *The Data: spontaneous speech*

An interesting peculiarity of Lorenzo’s speech regards his negation system. Firstly, it must be noted that his use of English Negative Polarity Items (NPIs) was highly adult-like:

- |     |                                 |       |
|-----|---------------------------------|-------|
| 61. | a. I don't need any             | (3;5) |
|     | b. The cows didn't eat anything | (3;5) |
|     | c. I didn't found anything      | (3;5) |
|     | d. I didn't say anything        | (3;6) |

Examples such as (61b) and (61c) in particular indicate that Lorenzo was using the NPI structure productively, rather than just imitating adult use. In fact, Lorenzo’s NPI system seems to have been even more advanced than that of some of his monolingual peers (from the Brown corpus, Brown 1973):

- |     |  |             |
|-----|--|-------------|
| 62. | a. I don't like some                   | (Adam 3;2)  |
|     | b. I won't have some more              | (Adam 3;5)  |
|     | c. He didn't ha(ve) some(thing) to eat | (Sarah 3;3) |
|     | d. You didn't get some water in there  | (Sarah 3;7) |

However, this is true only as far as NPIs are concerned. By the same age, Adam and Sarah had already begun using Negative Quantifiers<sup>21</sup>:

63. Adam (2;11)

\*CHI: telephone # yeah .

\*CHI: no # nobody .

\*MOT: nobody's there ?

64. Sarah (2;9)

GLO: Sarah # what do you have on your feet ?

\*CHI: nothing .

In contrast, Lorenzo seemed exceptionally reluctant to use Negative Quantifiers (NQs). When faced with a wh-object question, he would consistently use NPI structures as in (61d). This is in itself surprising. In an experiment designed to test the development of NPIs in monolingual English children, Sun Song (1999) found that children between the ages of 3;1 and 5;6 strongly preferred object NQs over NPI structures when compared to the adult controls. The adult's responses showed a clear bias towards NPI use. NPI structures made up 83% of the adult responses to questions of the type "what is the boy touching?" whilst NQ structures were used only 14% of the time. The children's responses showed a bias in the opposite direction. NQs were used by children 64% of the time while NPI structures made up a mere 16% of the responses<sup>22</sup>.

However, Lorenzo's problem with NQs became obvious when he was asked wh-subject questions and thus could not opt for an NPI structure. In these cases he would respond in a non-adult manner:

65. OBS: Who helped you?

Lorenzo: Only me (3;5)

66. OBS: Who is going to drive you there?

Lorenzo: Only me (3;5)

67. OBS: Who went with you?

Lorenzo: My byself (i.e. by myself) (3;6)

The potentially formulaic nature of this utterance together with the fact that it was the only NQ uttered by Lorenzo in over four weeks, indicates that these elements may not have been part of the child's competence<sup>23</sup>. Indeed, results from a repetition task provide some support in favour of this view.

It has often been suggested that there is a systematic relationship between the structures a child hears on the one hand, and the structures s/he produces as a manifestation of his/her own competence on the other<sup>24</sup>. Following this view, repetition tasks allow us a glimpse of how the child's system diverges from the adult target. With this in mind, I told Lorenzo that we would be learning a song. I then began singing "For he is a jolly good fellow" asking him to repeat after me. Lorenzo's remarkable response is given below:

This example provides further support in favour of the view that Lorenzo's grammar was unable to produce NQ structures. Although he appears to have understood that the NP subject involved "zero people", he used a structure involving a universal quantifier with sentential negation, suggesting that the lexical instantiation of the  $\forall\neg$  sequence was not part of his knowledge. A similar example, though involving a different surface scope, was uttered spontaneously on arriving at the local playground. Noting that there were no other children there, Lorenzo looked annoyed and uttered (70):

While the response in (69) involves a universal scoping over negation, the sequence in (70) is intended with negation taking scope over the existential subject. Given that this is not the surface scope of (70), it seems that Lorenzo had to go to some length as a way of avoiding NO use. This presumably involved

lowering/restructuring of the quantifier in order to achieve the intended  $\neg\exists$  interpretation.

This contrasts sharply with the performance of monolingual English children (cf. 63), thus suggesting that Lorenzo's difficulties with English NQs may be due to interference from Italian. However, unlike the cases discussed so far, this linguistic behaviour is not immediately traceable to the child's other L1. Most notably, it is not the case that Lorenzo produced NQs in a manner reminiscent of the distribution of Italian n-words<sup>25</sup>, as he failed to produce NQs altogether.

Perhaps the most obvious way of approaching this is to see whether Italian n-words usually develop later than English NQs, thus opening up the possibility that, somehow, they might have slowed down development in the other language too. This seems to be the assumption made by Paradis and Genesee (1996) when defining delay as a type of cross-language interference. Simply put, this is the idea that the development of a structure in L1a might be delayed if the same structure involves slower development in L1b. However, this does not seem to be a viable solution. A search of various Italian corpora available on the CHILDES database (MacWhinney and Snow 1990) revealed that Italian children begin to use n-words roughly at the same age as their English peers develop NQs, if not earlier. The following examples are taken from the Calambrone, Tonelli and Antelmi corpus respectively:

#### 71. *Early Occurrences of n-words in Italian Children*

a. qua apre      la porta e non c'è nessuno      (Marco 2;3)  
 here open.3SG the door and not there-is n-word  
 'Here he opens the door and there is nobody'

b. \*MOT:      il sedano e poi ?  
                 the celery and then  
       \*ELI:      e poi niente .      (Elisa 2;1)  
                 and then n-word

c. da non farli      mangiare a nessuno      (Camilla 3;4)  
 of not make-them eat.INF at n-word  
 'so that nobody eats them'

This is also reflected in the speech of Lorenzo who did use Italian n-words, though sporadically:

72. OBS: E tu cosa hai fatto?  
and you what have.2sg done  
'And what did you do?

Lorenzo: Niente (3;5)  
n-word

Consequently, it seems that we cannot appeal to an effect of the kind suggested by Paradis and Genesee (1996) whereby the English system might have developed later 'in sympathy' with the Italian counterpart. In the next section I will try to tackle the problem from a theoretical perspective. I will first attempt to establish what properties the two negation systems might comprise and whether Cascade-updating may have been responsible for the interaction of these in a manner that lead to the linguistic behaviour exemplified in (65)–(70).

#### *4.1.3 Theoretical Considerations*

In order to see how the Cascade-updating model might help us understand the effects underlying examples (65)–(70), we must rely on existing theoretical assumptions in relation to the negative systems of English and Italian. Unfortunately, however, although there seems to be general agreement in relation to the English system<sup>26</sup>, the nature of n-words is very much debated. Moreover, even those theories that have made some headway within this area found the Italian system to be particularly problematic. This makes the acquisitionist's task much harder. Nevertheless, the data presented in (65)–(70) are very interesting and, as far as I know, nothing of the sort has ever been reported in the literature. I will therefore attempt to address this peculiarity of Lorenzo's grammar in the following fashion. Firstly, I will give a brief overview of the relevant literature on negation and n-words, with particular focus on Italian. I will then discuss how Lorenzo's data might be explained by assuming the Cascade-updating model in conjunction with some empirically motivated assumptions.

#### **4.2 Negative Concord**

Negative Concord (henceforth NC) is the phenomenon whereby structures are interpreted as having a single negation although they seem to contain more than

one negative element. This is found in many of the world's languages. One example is given by the Italian<sup>27</sup> element *nessuno*. When used in isolation, this element appears to be negative, as in (73):

73. Q: Chi avete visto?  
who have.2PL seen  
'Who have you seen?'

A: Nessuno.  
Nobody

When occurring within a sentence, however, sentential negation must also be present:

74. \*(non) ho visto nessuno  
not have.1sg seen n-word  
'I haven't seen anyone'

This differs sharply with NQs (as used in Standard British English) which are always unambiguously negative. Their negativity is evident in cases where they co-occur with sentential negation to yield a double negation reading. The English examples below illustrate this point:

75. a. I saw nobody (I didn't see anybody)  
b. I didn't see nobody (I saw somebody)

To complicate matters further, Italian n-words disallow sentential negation when they appear pre-verbally<sup>28</sup> (underlining indicates stress):

76. a. Nessuno (\*non) ha visto Mario  
n-word not have.3sg seen Mario  
'Nobody has seen Mario'
- b. Con nessuno (\*non) ho parlato  
with n-word not have.1sg spoken  
'I spoke to nobody'



This negative concord 'puzzle' has received much attention in the literature. Essentially, the question that has been repeatedly asked is whether or not n-words are inherently negative, though the situation remains pretty much unresolved.

In the next sections I will briefly review the most influential analyses and mention some of the problems that have been raised.

#### 4.2.1 *N-words as negative*

For those who proposed that n-words are inherently negative (see for example Zanuttini 1991), the challenge is to find an explanation for the requirement exemplified in (74). In her extensive work on negation, Zanuttini (1991, 1995, see also Haegeman and Zanuttini 1991) proposes that n-words are in fact negative quantifiers which undergo what she calls an 'absorption' rule. This is a rule that takes multiple occurrences of negative quantifiers and collapses them into a single instance:

##### 77. Negative absorption

$$[\neg x] [\neg y] [\neg z] \rightarrow [\neg x, y, z]$$

This analysis has raised a number of theoretical as well as empirical issues.

Firstly, the process in (77) has been argued to be against the spirit of compositionality, as its central role is to somewhat invalidate the meaning contributed by some of the composing lexical items (May 1989). Moreover, although negative absorption provides a formalisation of the NC phenomenon, its status as a rule is called into question when we consider that it does not seem to be derivable from any general principle of the grammar. It does not really help us understand NC but it merely formalises its anomalous status (cf. Giannakidou 2006). The question also arises as to what determines whether negative absorption applies in any particular language.

One way of dealing with this question would be to assume that negative absorption is parameterised (i.e. it applies in Italian but not in English). However, this does not seem to improve the situation. Since the beginning of the Minimalist Program (Chomsky 1993), syntactic parameters have been abandoned in favour of feature strength/lexical variation analyses, a paradigm change which no longer allows the postulation of parameterised rules like that in (77).

As to the nature of negative absorption, the problem might disappear if we simply opt for a different notational variant. This is the position advocated by, for example, Brown (1999) and Watanabe (2004) who propose that the negative

absorption rule can be reformulated in terms of feature checking (Chomsky 1995, 2000). The main goal of these proposals is to derive the NC phenomenon from principles which find independent motivation within the Minimalist Program. For example, Brown (1999) suggests that the negative feature carried by the n-word is uninterpretable and must therefore be deleted by feature checking against a negative head carrying an interpretable [neg] feature. Sentences like (74) are thus correctly predicted to involve only one semantic negation despite requiring the presence of a negative marker (the negative head in Brown's account). A similar proposal is put forth by Watanabe (2004), although he argues that the underlying trigger for the checking operation in NC structures is an uninterpretable [focus] feature rather than a [neg] feature.

Nevertheless, although these analyses may offer an improvement in virtue of their compatibility with standard Minimalist assumptions, they face one important empirical problem that has been raised by their predecessor, namely the incompatibility with cases like (76) above. Even if we were to agree that negative absorption/feature uninterpretability is active in Italian, further stipulations need to be made in order to prevent these from applying when the n-word appears pre-verbally (cf. 76).

A number of researchers have argued for the alternative analysis, namely that n-words do not involve a negative element (Giannakidou 1998, 2000, 2006; Laka 1990; Ladusaw 1992; Rizzi 1982 among others). Of course, examples such as (74) are unproblematic for these proposals, their challenge being with regard to the data in (73) and (76) instead. Work within this tradition can be roughly separated into three categories; existential, universal and indefinite analyses. I will discuss each of these in turn.

#### 4.2.2 *N-words as Existentials*

Perhaps the earliest proposal to consider n-words as non-negative is Rizzi's (1982). Rizzi's view is that n-words are existential quantifiers which can incorporate negation when c-commanded by clausal negation. This analysis is based on the application of Quantifier Raising which is responsible for turning structures like (74) into the LF in (78).

78. [non + nessuno<sub>i</sub> [ho visto e<sub>i</sub> ]]

In order to cover the data in (76), however, a further rule needs to be stipulated. Rizzi (1982) suggests that these quantifiers are sensitive to a rule of negative

incorporation which associates the feature [+ neg] to them when they occur clause initially. This feature association transforms the existential quantifier into a negated existential (adapted from Rizzi, 1982: 124):

79. Q -> [+ neg] when c-commanded by VP

The most obvious question this approach raises is where the alleged [+neg] feature comes from. Like negative absorption, the rule in (79) seems to be against the spirit of compositionality as it introduces a property which is not associated with any of the lexical components<sup>29</sup>. It also finds no support beyond the data it is meant to capture, thus appearing to be a mere reformulation of the problem.

#### 4.2.3 N-words as indefinites

Ladusaw (1992, 1994) proposes that n-words are indefinites with descriptive content but no quantificational or referential force, (what he calls 'Hemian' indefinites, presumably referring to the work of Heim (1982)). He then suggests that n-words differ from the NPI indefinites found in languages like English with regard to their licensing requirements. N-words, Ladusaw argues, require licensing at the semantic level by being in the scope of a negative operator, which he calls the roof. NPIs also need a roof, but in addition they require that the roof be in a c-commanding position in overt syntax.

A typical example showing that scope and c-command do not always coincide is given below (Ladusaw 1992: 10):

80. A train didn't arrive for three hours

The sentence in (80) allows a reading where "for three hours, it has not been the case that *any* train arrived". Under this reading, negation takes scope over the indefinite "a train" despite the fact that it does not c-command it. However, it is well known that NPIs cannot be licensed in the same position:

81. \*Anybody didn't arrive

Based on the idea that all indefinites need a licensing roof and in view of the contrast between (80) and (81) above, Ladusaw concludes that c-command is an additional requirement peculiar to NPIs and not to indefinites in general<sup>30</sup>. In other

words, NPIs are subject to a stricter licensing condition requiring that they be c-commanded by as well as in the scope of a negative operator.

Besides offering an explanation for the contrast between (80) and (81) above, this claim also predicts that there should be cases where NPIs fail to be licensed despite the presence of a c-commanding antecedent. The existence of scope constraints (Linebarger 1987) suggests that this is indeed the case. The contrast between (82a) and (82b) below shows that c-command is not a sufficient condition for NPI licensing.

82. a. Meg didn't read every book to a student  
b. Meg didn't read every book to any student

The sentence in (82a) has a reading in which "it is not the case that every book got read". This is because the indefinite "a student" falls in the scope of the quantificational NP "every book" which, in turn, is in the scope of negation yielding the sequence  $\neg\forall\exists$ . However, this reading is not available in (82b).

This is due to the presence of an intervening quantificational NP that blocks the negative operator from taking scope over "any" (Linebarger, 1987). As a result, the negative operator fails to license the NPI<sup>31</sup> even though it c-commands it.

The view that n-words are indefinites is also empirically well motivated. In particular, Ladusaw draws attention to their distribution in non-negative sentences. Unlike negative elements (e.g. the quantifiers in 75), when n-words appear in some non negative contexts they yield a non-negative meaning (Rizzi 1982; Acquaviva 1999; Mathieu 2001):

83. a. Ha telefonato nessuno?  
has telephoned n-word  
'Has anyone phoned?'
- b. Chi ha detto niente?  
who has said n-word  
'Who said anything?'
- c. Dubito che ne sappia niente  
doubt.1SG that of-it know.3SG.SUBJ n-word  
'I doubt that he'd know anything about it'

- d. Sono        sorpreso che conosca        nessuno  
 be.1SG       surprised that he know.3SG.SUBJ n-word  
 'I am surprised that he knows anyone.'
- e. Tutti quelli che       sanno       niente a proposito di  
 all   those who       know.3PL n-word about  
 cio` sanno       che è pericoloso  
 this know.3PL that is dangerous.  
 'Everyone who knows anything about this knows it's dangerous'
- f. me ne       sono       andato senza       dire niente  
 I   of-it be.1SG gone       without say n-word  
 'I left without saying anything'.

The facts in (83) can be straightforwardly accounted for if we follow Ladusaw in assuming that n-words are non-negative indefinites<sup>32</sup>. As they are not negative, they can occur in a number of non-negative environments. Moreover, if they do not have referential force we expect them to lend themselves to different interpretations depending on the nature of their licenser.

However, note that an indefiniteness analysis is not the only option compatible with the data in (83). As pointed out by Rizzi (1982) and more recently by Giannakidou (2006), these data could easily be accommodated by analyses that take n-words to be polarity sensitive existentials. However, I will suggest that this difference is trivial as far as acquisition is concerned. I will return to this point in section 4.3.

Returning to Ladusaw's proposal, although it is probably one of the most successful to date, it still faces some of the issues that have troubled all the other non-negative views of n-words, namely the data in (73) and (76). Ladusaw's view is that these are instances of 'self-licensing' whereby the insertion of an n-word in the syntax triggers the building of a NegP. The NegP, in turn, licenses the n-word itself. In other words, n-words do have a [neg] feature of some kind (hence the triggering of a NegP) but they are not semantically negative. It is not obvious whether this differs from an account which postulates empty negative operators and it is not clear how the postulation of this [neg] feature impacts on the analysis Ladusaw offers for the cases in (83). Moreover, this move still leaves the question of what underlies the subject/object asymmetry expressed in (74) and (76). If n-words can license themselves, why must sentential negation be present in (74)?

Perhaps a more serious problem is posited by the prediction that the distribution of NPIs should be a subset of the distribution of n-words. This follows from the assumption that c-command and scope are not interdependent; while a c-commanding relation entails scope, a scope relation does not entail c-command. Nevertheless, Acquaviva (1999) discusses a number of environments where NPIs - but not n-words - are successfully licensed, suggesting that these two groups of elements are more accurately described as being in an intersection rather than a subset relation. These involve the following (adapted from Acquaviva 1999):

84. *Inside a subject*

- a. \*non<sub>i</sub> possiamo affermare che [nessun<sub>i</sub> caso]  
not can.1PL assert that n-word case  
sia risolto  
is solved  
‘we can’t assert that any case has really been solved’
- b. we can’t claim that any case has really been solved

Note that the ungrammaticality of (84a) cannot be due to locality constraints since the n-word can be licensed in object position:

- 85. non<sub>i</sub> possiamo affermare che la polizia abbia davvero risolto  
not can.1PL assert that the police has really solved  
[nessun<sub>i</sub> caso]  
n-word case  
‘we can’t assert that the police have really solved any case’

86. *Within Adjuncts*

- a. \*non<sub>i</sub> ho reclamato [a causa di nessun<sub>i</sub> problema tecnico]  
not have.1SG complained because of n-word problem technical
- b. I didn’t complain because of any technical fault

87. *Wh-island*

- a. \*non<sub>i</sub> mi chiedo [se tu abbia detto niente<sub>i</sub>]  
not me ask.1SG whether you have said n-word
- b. I am not wondering whether you said anything

88. *Relative / Complex NP*

- a. \*non<sub>i</sub> conosco [una ragazza che ha comprato niente<sub>i</sub>]  
 not know.1SG a girl that has bought n-word  
 b. I don't know (of) a girl who bought anything

These examples seem to force an analysis whereby n-words and NPIs do not enter a subset relation. As Acquaviva (1999) points out, the respective licensing environments are a lot less similar than they might at first seem.

4.2.4 *N-words as Universals*

Giannakidou (1998, 2000, 2006) provides an extensive discussion of various negation systems and concludes that negative concord is not a uniform phenomenon, a position which may turn out to be rather fruitful. As far as Greek is concerned, she argues that n-words are universal quantifiers. This, together with the assumption that Quantifier Raising applies at LF, seems to make for a rather simple solution for those language that – like Greek – always require sentential negation to license n-words. Compare (89) with (74) and (76) above (from Giannakidou 2006):

89. a. O Petros \*(dhen) idhe kanenan. Greek  
 the Peter not saw.3sg n-person  
 'Peter did not see anyone.'
- b. Kanenas \*(dhen) idhe ton Petro.  
 n-person not saw.3sg the Peter  
 'Nobody can see Peter.'

Giannakidou concludes that Greek n-words are polarity sensitive universals which need negation in order to be licensed. As far as this analysis is concerned, pre-verbal n-words are the unproblematic case. If n-words are universals, then it follows that they need to precede negation in order to yield the correct interpretation, namely  $\forall\neg$ . In post-verbal cases, on the other hand, QR is called upon as a means to achieving the desired interpretation. Details aside, the example in (89a) will have the LF in (90):

90. [kanenan<sub>i</sub> [o Petros dhen idhe t<sub>i</sub>]]

This leads to the correct scope relation, namely  $\forall\neg$  and seems to capture the Greek data fairly straightforwardly. Note that the assumption underlying the analysis in (90) is empirically well motivated. For example, it is generally assumed that QR cannot cross a clause boundary<sup>33</sup> (Reinhart 1997). Greek presents a potential exception to this. Firstly, it must be noted that Greek has three types of complement clauses, namely factive, non-factive and subjunctive which are introduced by the complementisers *pu*, *oti* and *na* respectively (Giannakidou 2006). The important observation in relation to these is that the scope of a universal quantifier is clause-bound only as far as *pu* and *oti* clauses are concerned, as shown in (91)-(93), from Giannakidou (2006:12-13).

91. a. Kapjos kathijitis frondise kathe fititis s'afti ti lista na  
       some professor ensured.3SG every student in-this the list SUBJ  
       vri dhulja  
       find.3SG job  
       'Some professor made sure that every student on this list will find a  
       job'  
       b.  $\exists > \forall$   
       c.  $\forall > \exists$
92. a. Kapjos fititis lipithike pu kathe kathijitis tis sxolis  
       some student was-sorry.3SG that every professor the department  
       apolithike  
       got-fired.3SG  
       'Some student regrets that every professor in the department got  
       fired'  
       b.  $\exists > \forall$   
       c.  $*\forall > \exists$
93. a. Kapjos fititis ipe oti kathe kathijitis tis sxolis  
       some student said.3SG that every professor the department  
       apolithike  
       got-fired.3SG  
       'Some student said that every professor in the department got fired'  
       b.  $\exists > \forall$   
       c.  $*\forall > \exists$



As the analysis given in (90) claims that Greek n-words are polarity sensitive universals, the prediction that follows is that n-words too should be sensitive to clause boundaries in the same way as universal quantifiers. In other words, licensing of a Greek n-word should be clause bound unless the n-word is contained within a *na* complement. This appears to be the case (adapted from Giannakidou 2006:11):

94. O Pavlos dhen theli na dhi kanenan  
the Paul not want.3SG SUBJ see.3SG n-word  
'Paul doesn't want to see anybody'

95. \*Dhen lipame pu pligosa kanenan  
not be-sorry.1SG that hurt.1SG n-word  
'I don't regret that I hurt anybody'

96. \*O Pavlos dhen ipe oti idhe kanenan  
the Paul not said.3SG that saw.3SG n-word  
'Paul didn't say he saw anybody'

These data suggest that the domain within which Greek n-words are licensed corresponds to the QR domain, thus supporting the view that Greek n-words are polarity sensitive universals. What sets this analysis apart from the ones we have seen so far is that it relies on an assumption, namely the application of QR, which has been argued to be necessary for independent reasons (see for example May 1985; Diesing 1992).

As to one-word uses (cf. 73), Giannakidou proposes that these are elliptical structures. Therefore, the answer in (73) would have the structure in (97):

97. [~~non-abbiamo visto~~] nessuno  
not have.1PL seen n-word

Although I am not particularly sympathetic towards the ellipsis view, it may be a small price to pay for a comprehensive analysis of what Giannakidou calls strict NC languages (Greek, Polish, Hungarian, Japanese, and most of the Slavic languages).

As to Romance languages, Giannakidou argues that n-words may be existential – rather than universal – quantifiers. This position is actually very similar to that of Ladusaw (1992) (cf. section 4.2.3). In both views, post-verbal n-words

yield a negative meaning due to their scope relation with a licensor while pre-verbal n-words are the special case. The two views differ only in so far as Ladusaw (1992) appeals to self-licensing while Giannakidou calls upon the idea of a neg trigger supposedly associated with n-words in a high sentential position (an idea that dates back to Zanuttini 1991).

In relation to this discrepancy between pre- and post-verbal n-words, we may want to take seriously the intuition that focus is somehow involved in the licensing of (some) NC structures (Giannakidou 1998, 2000; Tsimpli and Roussou 1996; Watanabe 2004). Although work so far has only related focus to Greek and Japanese NC<sup>34</sup>, the Italian data in (74) and (76) are indicative of an interesting pattern; whenever the n-words appear without negation they bear sentential stress. In fact, this seems to be a necessary requirement. Compare the following examples from English and Italian:

98. a. Nobody spoke  
b. Nobody spoke
99. a. Nessuno ha parlato  
n-word has spoken  
b. \*Nessuno ha parlato  
n-word has spoken

Of course, no such requirement exists for regular DPs, as shown in (100).

100. a. Mario ha parlato  
Mario has spoken
- b. Mario ha parlato  
Mario has spoken

Assuming that focus can license n-words would explain the facts in (99) as well as their contrast with (98): the English NQs do not need any licensing and are therefore grammatical regardless of sentential stress, although a slightly different pragmatic implication may be conveyed. The same applies to the cases in (100).

The question arises as to why focus-licensing is in fact obligatory in pre-verbal cases (cf. 76a-b). However, there may be independent reasons for this. The structure [nessuno...not] would need to undergo quantifier lowering in order to

achieve the correct scope relation, namely  $\neg\exists$ . If, as Chomsky (1995) suggests, LF-movement only applies as a last resort, there may be pressures to avoid it when an alternative license is available. If we accept that focus can license n-words, then it follows that in preverbal cases focus-licensing is not merely possible but obligatory.

Inevitably, this raises a question in relation to why the same focus licensing is not available in the post-verbal cases. This too may be due to independent reasons. In structures like (74) above, the negative element is located somewhere between the C-head and IP. If we follow Rizzi (1997, 2001) in assuming that focus Operators are situated somewhere in the left periphery of the clause (see also Acquaviva 1999), then an answer can be found by appealing to some minimality condition, such as the Minimal Link Condition of Chomsky (1995). We can then correctly rule out focus licensing for post-verbal n-words given that a closer licenser is available in the form of the negative marker *non*.

In Greek, on the other hand, the only possible licenser for n-words is the negative marker, which is always obligatorily present. This also seems to follow from the assumption that Greek n-words are universals. Preverbal n-words in Greek do not need to undergo LF lowering since the order in which they surface involves the correct scope relation, namely  $\forall\neg$ . Consequently, there are no economy considerations that would force Greek n-words to find an alternative licenser.

In the next section I will argue that if the split between strict and non-strict NC suggested by Giannakidou (2000) is correct, Lorenzo's case may find an explanation even if we do not commit to any particular analysis of pre-verbal n-words in Romance. This is because there is evidence that – whatever property is responsible for the behaviour of pre-verbal n-words – it may not be relevant to early development.

### **4.3 N-words: The Children's View**

#### *4.3.1 Introduction*

In the previous sections we have seen that there may be at least three ways in which languages can establish negation: Negative Quantifiers (e.g. English, cf. 75a), polarity sensitive Universals (e.g. Greek, cf. 89) and polarity sensitive existentials (e.g. Italian, cf. 76). Although most of the literature on NC has been concerned with which of these options applies in which language, there seems to be little controversy with regard to the idea that negation can be expressed in more than one way. As has been pointed out countless times, the following are truth conditionally equivalent, though logically different:

101. *Logical representation of negative statements*

- a.  $\forall x [P(x) \rightarrow \neg Q(x)]$       Universal scoping over negation  
 b.  $\neg \exists x [P(x) \wedge Q(x)]$       Negation scoping over existential

Taking this as a point of departure, we must allow for at least two more possibilities since each of the logical relations in (101) can be linguistically expressed in more than way, namely syntactically or lexically:

102. *Possible instantiations of  $\forall \neg$* 

- a. **Kanenas dhen** idhe      ton Petro.      Syntactic<sup>35</sup>  
     n-word   not   saw.3SG   the Peter  
     'Nobody can see Peter.'  
 b. **Nobody** saw Peter      Lexical

103. *Possible instantiations of  $\neg \exists$* 

- a. I did **not** see **anybody**      Syntactic  
 b. **Nessuno** ha   parlato      Lexical<sup>36</sup>  
     n-word   has   spoken

If we assume, rather uncontroversially, that the relations in (101) are the basic building blocks of negative systems and (102)-(103) their possible implementations, we may put on hold concerns about the details of negative systems and shift our focus towards the constraints that these basic structures may put on child language. In particular, we can focus on the input that a child is faced with when exposed to a given negation system and on trying to establish which basic relation might be intimated by such input. The benefit of this shift is twofold. First, it might help us gain a better understanding of what lexical properties the child initially assumes for a given language, regardless of what the target setting might be. Secondly, we might shed some light on why children make the errors they do when acquiring negation systems. Given its part in the Cascade-updating process, it is the former which will be given most attention here.

In the next sections I will consider the options in (102) and (103) above in relation to the child's input and argue that while the English input might be compatible with the interpretations in (102a) and (102b), the Italian input suggest an interpretation in terms of (103a).

### 4.3.2 Italian Input

As the examples in (74) and (76) showed, n-words can appear both pre- and post-verbally. However, a search of Italian corpora revealed that n-words are much more common in object than in subject position. A corpus of the BADIP database (a database of spoken Italian) was searched for all instances of *niente* as well as *nessuno* (masc), *nessuna* (fem) and their allomorph *nessun*. In an attempt to only strictly consider the spoken variety, public speeches were excluded from the search, as these are effectively a version of written Italian. A total of 459 tokens were found amongst the 489178 words recorded in the database. Of these, all the occurrences of *nessuno* (and its variants) were included in the count. As to *niente*, 99 occurrences were excluded from the count as these represented interjection uses. These are some kind of parentheticals which resemble single-word uses, whereby the n-word is often preceded by hesitation. The examples in (104) and (105) from the BADIP corpus exemplify this type of use:

104. Sta all' ambasciata a Lussemburgo e e niente lì la scuola  
 stays at-the embassy at Luxemburg and and n-word there the school  
 finisce a diciott' anni  
 ends at eighteen years  
 'He works for the embassy in Luxemburg and.. and.. well, there you go to  
 school till you are eighteen.'
105. Dopo di che eh niente dopo di che io fino a mezzanotte sono stata  
 after of that erm n-word after of that I till at midnight be.1SG stayed  
 'and after that, well after that nothing I stayed till midnight '

Although this use appears to be very common (almost a third of the tokens found involved this expletive use) it does not seem to be a potential source of information to the child as far as the negative system is concerned.

One-word answers were also excluded (85 tokens) since these provide no clue to the child as to the status of the elements at issue<sup>37</sup>. The remaining 274 tokens were therefore representative of the input the Italian child must rely on when acquiring the negative system.

The results indicate that there is a strong imbalance in the distribution of n-words when comparing the subject and the object position:

106. *Distribution of n-words in spoken Italian corpus*

<b>subject</b>	<b>object</b>	<b>other<sup>38</sup></b>
42/274 (15.3%)	207/274 (75.5%)	25/274 (9.1%)

Moreover, it must be noted that while a structure such as that in (74) is very common, Italian also allows subject n-words to surface post-verbally, as in (107).

107. a. Mario non lo ha visto nessuno  
 Mario not him-CL has seen n-word  
 'Nobody has seen Mario'

- b. Non ha parlato nessuno  
 Not has spoken n-words  
 'Nobody has spoken'

These structures seem to be very common among Italian speakers as they made up about half the amount of subject n-words in the BADIP corpus. Once we factor this in, we find that, in Italian, n-words are remarkably uncommon in pre-verbal position:

108. *Distribution of pre- and post-verbal n-words in spoken Italian corpus*

<b>pre-verbal</b>	<b>post-verbal (subjects + objects)</b>	<b>other</b>
22/274 (8%)	227/274 (82.8%)	25/274 (9.1%)

Given these findings, it appears that the Italian child receives plenty of evidence suggesting that n-words can occur in the scope of negation, potentially indicating that the n-word may be an existential quantifier, as in (101b). However, this is not the only option. Given that QR could be involved, post-verbal n-words are also compatible with the structure in (101a). This leaves the child with a choice between the two analyses in (109a) and (109b):

109. a. [non ... nessuno] (existential: in situ analysis)  
 b. [nessuno<sub>i</sub> [non t<sub>i</sub>]] (universal: QR analysis)

Importantly, however, learnability considerations suggest that the child should choose (109a) first. If the child erroneously assumes the structure in (109a) when exposed to a language that employs (109b), s/he can always retreat once s/he meets preverbal occurrences of the n-words scoping over negation (this would probably be the case if the target language is Greek, cf. (102a)). However, if the Italian child were to assume (109b) based on its compatibility with the post-verbal instances, retreat would be much more problematic since it would require negative evidence with regard to the judgements in (76a).

Moreover, the results in (108) indicate that the child is exposed to a high frequency of post-verbal subjects, providing him/her with positive evidence that n-words must systematically follow negation, regardless of their relation with the predicate. This seems to reinforce the child's initial assumption (i.e. 109a) thereby encouraging him/her to hang on to it for a rather protracted period of time.

Although the child will eventually be exposed to more pre-verbal occurrences and will thus be able to adjust the grammar accordingly<sup>39</sup>, it seems evident that - at least at first - Italian children would assume that the target system corresponds to (109a) rather than (109b).

In sum, learnability considerations together with the high frequency of post-verbal n-words in the input, indicate that Italian children would at first assume that they are being exposed to the structure in (109a). As we shall see below, the situation is quite different for English children.

#### 4.3.3 English Input

In virtually any discussion about negation in English (not least the present one) it is taken for granted that negative quantifiers are widely used in subject as well as object position. Although the grammaticality of sentences like (75a) is indisputable, it turns out that the negative quantifier *nobody* is overwhelmingly uncommon in object position. A search of the National British Corpus revealed that the vast majority of uses of *nobody* involve the subject position. The distribution of the other NQ *nothing* on the other hand, involves mostly object uses. This means that, overall, the English child is exposed to a balanced distribution of NQs, though each NQ is far from being evenly distributed.

In order to match the Italian search as accurately as possible, 274 occurrences of NQs were randomly selected from the National British Corpus. These included 158 utterances involving *nobody* and 116 involving *nothing*, all of which represented non single-word uses. The results are as follows:

110. *Distribution of NQs in English corpus*

<b>pre-verbal</b>	<b>post-verbal</b>	<b>other</b>
161/274 (58.7%)	103/274 (37.5%)	10/274 (3.6%)

These data indicate that English children encounter plenty of evidence suggesting that NQs do not need a licensing antecedent. Consequently, English children will – from early on – only consider the two possibilities in (102), before eventually converging on the target, namely (102b).

As it turns out, some English children initially assume that they are being exposed to the sentential rather than the lexical option (cf. 102b). This was perhaps first reported by Klima and Bellugi (1966) and Bellugi (1967) in their longitudinal study of Adam and Sarah. This stage began with the first sentential uses of NQs and continued for a rather protracted period.

111. a. We don't make nothing (Adam 3;8)  
 b. I don't see nobody (Adam 3;10)  
 c. couldn't do nothin(g) (Sarah 3;5)  
 d. nobody didn't broke it (Sarah 4;5)

It is important to note that this does not reflect the variety spoken by the children's parents, as can be seen from the following utterances:

112. *Exchange between Adam and mother*

CHI: Mommy # I just don't want nothing in dere .

CHI: nothing in dere .

MOT: you don't want anything in there ?

No example involving (indirect) parental correction is found in Sarah's recordings. However, the negative sentences uttered by Sarah's parents consistently involve standard use of NPIs and NQs<sup>40</sup>.

It therefore seems that – at first – English children opt for the structure in (102a). This is perhaps less odd than it may at first seem if we consider that children might understand that sentences involving NQs are negative but they might not necessarily manage to process every component word. This, together with the fact that the input is replete with examples of sentence-initial NQs, suggests that children may have to start building their negation system based on



two bits of information (i) sentences involving NQs are negative (ii) NQs do not require a negative antecedent. Admittedly, these are compatible with (102a) as well as (102b) and the question remains as to why the latter is chosen. My inclination would be that the child follows some principle dictating that negative sentences have a Negative Phrase, essentially sentential negation. This would then cease to apply at some later stage. However, this is only a tentative explanation and I do not suggest that it be taken too seriously.

#### 4.3.4 Lexical Representations

In the sections above we have seen that while English children associate NQs with universal quantification (cf. 102a), Italian children are likely to take a different developmental route which leads to associating n-words with existentials. Importantly, neither of these options involves inherently negative elements, negation being supplied by the relevant sentential negative instead. Since the children do not use NQs/n-words in affirmative sentences, it seems reasonable to assume that they take these items to have some sort of polarity requirement which, following Giannakidou (1998) I will assume to be in the form of negative dependency<sup>41</sup>. Therefore, at the initial stages of negation development, n-words and NQs will be represented in the children's grammar as (113a) and (113b) respectively:

113.

- |    |                              |                   |     |                   |                               |
|----|------------------------------|-------------------|-----|-------------------|-------------------------------|
| a. | $\exists x, \text{dep(Neg)}$ | $\leftrightarrow$ | Det | $\leftrightarrow$ | $/\text{nessuno}/$            |
| b. | $\forall x, \text{dep(Neg)}$ | $\leftrightarrow$ | Det | $\leftrightarrow$ | $/\text{nəub\lambda d\iota}/$ |

We can now begin to see how Lorenzo's development may be accommodated within the Cascade-updating model. In the case of an English-Italian bilingual the two elements in (113) would - prior to the acquisition of quantification - share all their features and will therefore be perceived by the acquisition device as belonging to the same paradigm. Consequently, whichever specification the child happens to acquire first will affect the other. If it is Italian that acquires quantification first, then the application of Cascade-updating will result in transfer of the existential, as follows:

114.

$\exists x, \text{dep(Neg)}$	$\leftrightarrow$	Det	$\leftrightarrow$	/nessuno/
$\downarrow$				
$\exists x, \text{dep(Neg)}$	$\leftrightarrow$	Det	$\leftrightarrow$	/nəʊbʌdɪ/

Given the examples in (64) and (71), this outcome is not unlikely since it seems that the Italian elements appear in children's speech slightly earlier than their English counterparts.

However, this type of transfer differs from what we have seen so far in that the resulting element cannot easily be integrated into English grammar. Since English does not allow post-verbal subjects, the type of 'nobody' represented in (114) above cannot be produced as the subject of a clause. Of course, the child could still use this element in object position. However, in the case of Lorenzo we have seen that he had already mastered an alternative way of negating objects, namely NPIs (cf. 61). We therefore expect that the child would use NPIs whenever possible but lack an element that expresses negated quantification pre-verbally. This seems to be exactly what Lorenzo was doing.

If we follow Giannakidou's proposal in assuming that single-word answers involve elliptical structures, we could explain why Lorenzo would also fail to produce NQs as single answers. In order for him to be able to answer the questions in (65)-(67) grammatically, he would have needed to build a structure whereby the NQ was in subject position and then apply VP ellipsis. However, the element his grammar had acquired via Cascade-updating was not licensable in a structure of this type.

#### 4.4 Summary

In sum, Lorenzo's inability to form NQ structures in English can be given an explanation in terms of Cascade-updating. On the one hand, in section 4.2 we saw that the situation with regard to Negative Concord is still unresolved, despite decades of discussion. However, after looking at English and Italian child language (cf. section 4.3) it became possible to advance a hypothesis as to what properties children initially associate with NQs and n-words. According to this hypothesis, the two sets of elements differ minimally, thus being potential targets for the Cascade-updating mechanism. Finally, it was suggested that if application of Cascade-updating results in the transfer of the Italian setting to English, the resulting elements would be incompatible with pre-verbal uses, thus explaining why Lorenzo's English lacked *nobody/nothing* in these contexts. As to Lorenzo's

reluctance to use these same elements in post-verbal position, it was suggested that it may be due to the fact that he had already become proficient in the use of NPIs, and thus preferred these over the underdeveloped NQs.

## 5. Concluding Remarks

This chapter focused mainly on two case-studies involving two different language pairs: English-Italian and Cantonese-English. Both case-studies displayed signs indicating that (at least) one of the two L1s diverged from the path of development typically found in monolinguals. In both case-studies, this resulted in one of the two languages coming to host elements that behaved as those found in the other L1 and it was concluded that this was due to cross-language generalisation. The data were then given an explanation in terms of Cascade-updating. It was then argued that the same analysis is applicable to the case of Rosa and Guglielmo, two monolingual Italian children whose use of wh-elements diverged from that of their linguistic community.

Section 4 focused on the English-Italian case-study in more detail, though from a different perspective. Some data were presented showing that the subject of this case-study experienced a developmental delay whereby one of his two languages, namely English, did not seem to include elements which are usually present in the speech of monolingual children of the same age. In section 4.3 I suggested that this phenomenon differs from transfer effects only on the surface for - at the representational level - it can be maintained that delay is caused by cross-language generalisation just like the transfer effects discussed in sections 2 and 3. In other words, although the term 'delay' may be a useful shorthand description when referring to a scenario of the type presented in sections 4.1.1 and 4.1.2, it does not refer to any particular underlying process. In Lorenzo's case, we saw that the reason why the application of Cascade-updating surfaced as delay rather than as transfer effects was due to independent grammatical pressures. The elements that resulted from the application of Cascade-updating had two important characteristics: they were (i) ungrammatical pre-verbally (ii) in competition with a well-developed structure post-verbally. As a result, they were never included in Lorenzo's speech.

If the model I have presented so far is on the right track, the term 'transfer effects' is just a descriptive label for an overgeneralization process. Like in monolinguals, Cascade-updating indiscriminately adds newly acquired properties to a whole paradigm. Since the bilingual child has a larger lexicon, Cascade-updating

has a wider field of application and therefore, besides overgeneralization, transfer effects will also obtain. Crucially, the only difference between a monolingual and a bilingual child lies in the domain within which Cascade-updating applies. The fact that language production in bilingual children differs from that of monolinguals is then due to the very two aspects that they do not have in common, namely the input and the number of developing lexical sets. Moreover, assuming this model allows us to explain cases of delay without having to appeal to any additional machinery.

## NOTES

---

<sup>1</sup> In some Italian varieties, *che* does not necessarily require a complement and therefore the difference between the two elements seems much more subtle than our description would make out. However, the variety relevant to the case-study to be presented in section 2.1.2 does differentiate between *che* and *cosa* in the manner I discussed. I will therefore disregard regional differences since these do not seem to be relevant here.

<sup>2</sup> But see discussion on Guglielmo and Rosa in section 2.1.5 below.

<sup>3</sup> We cannot exclude the possibility that Guglielmo's and Rosa's development could have been influenced by other sources (e.g. television, playground) where a variety with a strong bias for *che* in pronoun uses may have been spoken (cf. footnote 1). However, even if this were the case, the question remains as to why the cues provided by these purported alternative sources should have been stronger than those provided by the parents.

<sup>4</sup> One of Guglielmo's utterances does involve *che* + a complement. However, this involves a different *che*:

- (i) *che naso lungo*.  
what nose long  
'what a long nose'

Firstly, as the gloss suggests, this is not a question use. Moreover, it involves a full DP (noun +Adj), a combination which is ungrammatical in the interrogative cases.

<sup>5</sup> The abbreviations used in the glosses are as follows: PERF = perfective, CLS = classifier, PRT = particle, ASP = aspect marker.

<sup>6</sup> The total number of children was in fact 31 but only 20 of them are relevant to our discussion as the remaining 11 were children with Specific Language Impairments.

<sup>7</sup> As the reader may have noticed, the two groups overlap in age: the oldest child from the younger group is older than the youngest child in the older group. This is due to the fact that Wong et al. (2004) needed one of the groups to be matched for mean age with their SLI subjects (see note 5). As we shall see in section 2.5.3, this division proved rather informative.

<sup>8</sup> No information was given by the authors about the specific child.

<sup>9</sup> On a couple of occasions, one particular child repeated the experimenter's utterance by adding the wh-element at the beginning. The experimenter's prompt and the child's response are given in (i) and (ii) respectively:

(i)    hungzai duk jan  
      bear    poke someone

(ii)    bingo hungzai duk jan  
      who    bear    poke someone

I will not consider these examples here.

<sup>10</sup> All the examples given by Wong et al involve the use of *bingo* ('who'). This is because the authors chose to use only one wh-word for the sake of consistency and it is not to be interpreted as related to any special status of this particular wh-word as far as the children's development is concerned.

<sup>11</sup> Note that the structure in (38b) also involves verb movement to C. The element *tung* – often translated as ‘with’ – is a type of verb known as a *coverb* rather than a preposition. See Wong et al (2004) for details.

<sup>12</sup> As movement to the C domain is not a feature of adult Cantonese, the question arises as to why this operation should be present in child Cantonese. Wong et al propose that this might be related to the fact that the structure in (38a) involves the same linear order as the adult structures for wh-subject questions. Consequently, movement in child Cantonese might be explained by assuming that the children associate interrogative sentences with wh/verb initial structures, possibly based on exposure to subject questions. This would lead to the instantiation of wh as well as verb movement when forming object questions.

Nevertheless, this raises the question of why subject questions in the input should provide such a strong cue.

<sup>13</sup> The use of the conditional (i.e. *could*) is probably more common in these contexts, but the examples I give are with the simple form *can* for the sake of consistency.

<sup>14</sup> But see Palmer (1990) for a different view.

<sup>15</sup> This is probably because it is a contradiction in terms. The Italian verb *riuscire* is comparable to the English verb *manage* as it refers to the subject’s ability to carry out an action. On the other hand – as we have seen – *potere* only allows a deontic reading, implying that there is some subject-external reason for why the action cannot be successfully carried out.

<sup>16</sup> I have asked four speakers from the Lombardy region what they thought about the sentence in (47) and they all confirmed that it sounded peculiar, unless it was intended with an epistemic interpretation. Note that television must be ruled out as a possible source of Standard Italian, and thus of uses of the type in (47), in Lorenzo’s home. This is because Lorenzo only had access to American channels, a decision made by his mother in an attempt to increase the amount of American-English input in an otherwise wholly Italian-speaking environment.

<sup>17</sup> Klinge (1993) actually argues that POTENTIALITY is a semantic field shared by a number of modals, among which are *can*, *may*, *must*, *will* and *shall*. I will not discuss this point here.

<sup>18</sup> Klinge (1993) relies on Relevance Theory (Sperber and Wilson 1986/1995) for the implementation of this process.

<sup>19</sup> A possible objection to this is that children have been reported to acquire deontic uses before epistemic uses, potentially indicating that the two are not intertwined. However, as argued by Papafragou (1998), this may be due to children's development of theory of mind rather than the actual semantic make-up of these elements.

<sup>20</sup> Müller (1998) also mentions the possibility of delay in BFLA. However, she appears to be concerned with a different concept altogether. The type of delay she seems to have in mind is that which might result after the surfacing of transfer effects. Her position can be summarised as follows. Transfer effects are a sign that the bilingual child has taken the wrong developmental route. Moreover, this route is not followed by monolingual children. Consequently, the bilingual child ends up in a situation whereby s/he will need to retreat from a non-adult setting, a situation which monolingual children do not have to deal with. This may result in the bilingual child attaining the adult setting later than his/her monolingual peers. A similar position is taken by Müller et al (2002). This differs quite radically from the type of delay I discuss in this section, which does not involve the emergence of transfer effects.

<sup>21</sup> These examples only focus on single-word answers. However, see section 4.3.3 for a brief discussion of sentential use in monolingual English children.

<sup>22</sup> 5% of the children's responses involved double negatives. No specific information is given about the remaining 13.8%, responses falling within this group are simply labelled as 'other'.

<sup>23</sup> Alternatively, it could be the case that some properties of these elements were indeed part of the child's competence but these were not sufficient for the elements to be integrated in his speech. This is the position I will take in section 4.3.4 below).

<sup>24</sup> This has been discussed by many researchers, perhaps mostly within phonology (Ingram 1974, Smith 1973, among many others), and it possibly dates back to the work of Poole (1934). More recently, Radford (1990) discussed this in relation to children's omission of functional categories. But see Smith (2003) for some criticism of this assumption.

<sup>25</sup> The term 'n-word' was first introduced by Laka (1990) as a neutral label for a number of elements, amongst which are the Romance *nessuno*, *ningu*, *personne* etc. I will use this term to refer to the Italian elements *nessuno*, *niente* whose behaviour differs from that of NQs (cf. section 4.2).

<sup>26</sup> In the last decade a debate has developed in relation to the alleged c-command requirement of Negative Polarity Items like *anybody/anything*. While some still maintain the GB-esque view that these require a c-commanding antecedent (for example Ladusaw 1992), others have argued that they only require to be scoped over by an appropriate element, c-command being a mere surface effect of this requirement (see for example de Swart 1998). As to elements like *nobody/nothing*, there is – to my knowledge – no disagreement as to their negative universal quantifier status.

<sup>27</sup> In what follows I will focus on Italian, however, see Brown (1999) for Russian, Espinal (2000) for Catalan and Spanish, Giannakidou (1998, 2000) for Greek.

<sup>28</sup> This is true of Italian and Spanish. Although the Catalan equivalent of (74) is grammatical, the absence of negation is not required in this language.

<sup>29</sup> Put in Minimalist terms, (79) is a blatant violation of the Inclusiveness condition. To be fair, however, this was introduced more than a decade after Rizzi's proposal.

<sup>30</sup> Although it must be noted that the picture is further complicated by the fact that NPIs contained in indefinites appear to be licensed in the same position:

- (i) A doctor of any standing did not appear to be available.



<sup>31</sup> The only NPI reading available is: "for every book it is not the case that there exists a student to whom Meg read those books" ( $\forall x: \text{book}(x) \neg (\exists y: \text{student}(y)) [\text{read}(\text{Meg}, x, y)]$ ). This also follows from Ladusaw's claim since this interpretation finds the NPI in the scope of negation:  $\forall \neg \exists$ .

<sup>32</sup> Although the examples in (83c) and (83f) do not seem to me to be evidence for indefinite status since both the preposition *senza* (without) and the verb *dubitare* (doubt) are arguably inherently negative.

<sup>33</sup> But see May (1989) and Fox (2000) for discussion on how this assumption may have to be slackened.

<sup>34</sup> In fact, Watanabe (2004) explicitly denies that focus is involved in Italian NC.

<sup>35</sup> This is also found in some English constructions such as (i) below:

(i)     **All** trains do **not** stop at Victoria                    ( $\forall \neg$ )

For the sake of consistency I will not discuss these and will keep focusing on n-words instead.

<sup>36</sup> Though English does not seem to make use of the lexical variant of (101b), this is what Rizzi's (1982) analysis associates *nessuno* with. This is also Ladusaw's view since the indefinite nature he attributes to Italian n-words requires them to be interpreted under existential closure. Consequently, Ladusaw's proposal also involves interpreting (103b) as  $\neg \exists$  (see also Mathieu 2001).

<sup>37</sup> Of course, single-answer examples will inform the child that the element at issue does not have the same requirements as English *anybody*. However, this fact alone does not help the child choose between the existential and the universal option (cf. (102) and (103)) since both of these are potentially compatible with a single-word use. The fact that NPIs have an additional requirement that restricts their distribution is not relevant to this point.

<sup>38</sup> This includes non-negative uses (5 tokens), 'mention' uses such as *Il signor nessuno* (Mr. Nobody) as well as uses which cannot be catalogued under any of the above.

<sup>39</sup> As our discussion in section 4.2 suggests, it is far from obvious what this adjustment would involve.

<sup>40</sup> Two points must be noted. Firstly, Sarah's mother did on one occasion repeat an NC structure after it had been uttered by Sarah:

- (i) CHI: took [//] couldn't do nothin(g) .  
MOT: she's right .  
MOT: couldn't do nothin(g) .

Although this might be reinforcing Sarah's NC system, the fact remains that the child's initial assumption is not relatable to parental speech.

The second point is that NC is rather common in colloquial and working class American-, as well as British-, English. It is therefore possible that Adam and Sarah may have been exposed to NC outside their home. For the purposes of our discussion, I will follow Klima and Bellugi (1966) and Bellugi (1967) in assuming that NC in Adam and Sarah's speech is due to developmental factors. However, this area is in urgent need of further and more thorough research that will control more closely for the potential presence of NC in non-parental input.

<sup>41</sup> In fact, Giannakidou proposes that Italian and Greek n-words are dependent on non-veridical, rather than negative operators. Although this may be the adult setting, children will initially assume a negative dependency until further evidence emerges (in Giannakidou's terms, children will begin with the assumption that they are acquiring a language with true NPIs).

## CHAPTER 3

---

# The Importance of Being a Paradigm

### 1 Introduction

The theory introduced in the first chapter makes two main claims. Firstly, it proposes that syntactic Transfer Effects in BFLA and overgeneralization in monolingual acquisition are manifestations of the same underlying mechanism. The second claim is with regard to the workings of the system responsible for these phenomena. This comprises an informationally monotonic mechanism, namely Cascade-updating, which interacts with a system of paradigm formation. While the former is responsible for propagating newly acquired information across parts of the lexicon, the latter defines the specific domains within which propagation takes place. Due to the nature of Cascade-updating, only items that share the relevant paradigmatic properties can affect each other and thus result in either overgeneralization or transfer effects. An important implication that follows from these claims is that both overgeneralization and transfer effects should only ever target items of the same paradigm. The aim of this chapter is to find support for these claims by showing that members of separate paradigms do not affect each other, despite the fact that they may serve a remarkably similar purpose on the surface.

The phenomenon that I will consider with regard to this claim is the Delay of Principle B effects which have been widely discussed in the acquisitional literature. In the sections that follow I will look at some experiments that have been carried out in relation to this phenomenon and I will argue that the results strengthen the tie between transfer effects and overgeneralization, thus corroborating the Cascade-updating model.

The chapter is organised as follows. Section 2 is an overview of the Delay of Principle B effects and of some theoretical approaches on the issue. Section 3 addresses some questions in relation to transfer effects in this domain and discusses some evidence from the literature. As to section 4, its aim is twofold. First, it presents the discrepancies that have been argued to exist between pronouns and clitics with regard to the delay of Principle B effects, and then it investigates the structural differences between these elements in an attempt to establish what underlies their divergent behaviour. Section 5 looks at what predictions emerge with regard to overgeneralisation across the pronominal paradigm, once we assume the model proposed in chapter 1. It then presents an experiment set up to test those predictions. Section 6 concludes the chapter.

#### 1.1 Some Current Issues

It has been argued, at least since Döpke 2000, that BFLA ought to be approached from a theoretical perspective so as to enable the development of specific hypotheses. Nevertheless, some researchers have repeatedly assumed that if the language system can produce transfer effects at all, then these should be found whenever two languages differ with regard to some property, the idea being that any difference automatically presents a case for transfer (Varlokosta and Dullart 2001, Zwanziger et. al. 2005, among others).

This assumption is discussed in some detail by Paradis and Genesee (1996), who recognise a number of possible effects that syntactic transfer should lead to. One of these is known as *acceleration*. On this view, a structure should appear in the speech of a bilingual earlier than in that of a monolingual if such a structure is clearly present in the child's other L1. In other words, the mastering of a structure in L1a should accelerate the development of the corresponding structure in L1b.

Although this may be a logical possibility, it is not compatible with the Cascade-updating model, as transfer effects are strictly limited to lexical items that belong to the same paradigm. According to this model, only items that share the relevant paradigmatic properties can affect each other and thus result in either overgeneralization or transfer effects. In other words, both overgeneralization and transfer effects are expected to only ever target items of the same paradigm.

In what follows I will look at some experimental evidence, both old and new, which seems to strengthen the tie between transfer effects and overgeneralization and, consequently, to provide evidence in support of the Cascade-updating model.

## 2. Delay of Principle B Effect

### 2.1 The Phenomenon

In the mid-nineties, much of the acquisitional literature was dedicated to children's development of Principles A and B of Binding Theory (Chomsky 1981, 1986a). Perhaps surprisingly, research revealed a divergence in children's development of these Principles. While children respect both the c-command and the locality requirements of Principle A by about age 5;6, they do not seem to obey Principle B until much later (see Chien and Wexler, 1990 and references cited there). In fact, children around age 5;6 often allow a pronoun to be co-referential with a local c-commanding antecedent, an interpretation that is ruled out in adult language.

In a particular study carried out by Chien & Wexler (1990), a group of 156 children ranging between age 2;6 and 6;6 were asked to act out what the puppet Kitty said in a variation of the popular "Simon says" game. The sentences uttered by the puppet included those in (1):

1.     a. Kitty says that Sarah should point to herself.  
       b. Kitty says that Sarah should point to her.

Children who were at least 6;0 years of age performed almost at adult level for sentences of the type in (1a) answering correctly in 90% of the cases, thus confirming the findings of previous works (see Chien and Wexler, 1990 for a detailed overview). However, the results decreased dramatically for sentences of type (1b). It was recorded that in 36% of the cases even children as old as 6;6 allowed the pronoun to co-refer with the locally c-commanding NP (in this case *Sarah*), allowing the sentence in (1b) to yield the interpretation *Sarah should point to Sarah*. For younger children the answers were at chance level.

This phenomenon, known as the Delay of Principle B effect (henceforth DPBE), has been found in children acquiring different languages (see, among others, Avrutin and Wexler 1992, for Russian; Sigurjonsdottir 1992, for Icelandic). These data are hard to reconcile with the claim that Principle B is a principle of Universal Grammar. On the one hand, Binding Principles seem to hold cross-linguistically and therefore would be expected to be universal and to present no difficulty for acquisition. On the other hand, children seem to find Principle B hard to master, indicating that some learning is involved.

An interesting solution has been developed by Grodzinsky and Reinhart (1993), who proposed a theory of competition at the syntax-pragmatics interface based on the distinction between *coreference* and *bound-variable* interpretation. An antecedent and a pronoun, they argue, can refer to the same entity via two separate processes, illustrated by the following (from Grodzinsky and Reinhart 1993:74):

2. Alfred thinks he is a great cook.
  - a. Alfred ( $\lambda x$  (x thinks x is a great cook))
  - b. Alfred<sub>i</sub> ( $\lambda x$  (x thinks he<sub>i</sub> is a great cook))

Although the interpretations that arise from the two readings are equivalent, there is a crucial difference at the conceptual level. In (2a) the pronoun is a variable syntactically bound by an R-expression while in (2b) the two expressions corefer by virtue of the fact that the discourse-referent chosen by the pronoun is the same as that introduced by the R-expression. Crucially, the latter reading does not involve a syntactic operation that relates the pronoun to an antecedent and therefore its interpretation is not regulated by Binding Theory. This raises a question with regard to the structures in (3b) and (3c). Since coindexation is ambiguous between coreference and bound variable interpretation, italics is adopted to represent the former:

3.
  - a. Lucie<sub>i</sub> adores her<sub>x</sub>
  - b. \*Lucie<sub>i</sub> adores her<sub>i</sub>
  - c. \**Lucie* adores *her*

Following standard Binding Theory, the ungrammaticality of (3b) does not present a problem as it is a violation of Principle B. On the other hand, coreference is not a syntactic operation and therefore the ill-formedness of (3c) cannot be captured syntactically. Also, note that it is not immediately obvious why coreference should be ruled out in (3c) as it is generally undisputed that it is not necessary for a pronoun to enter a syntactic dependency in order to be coreferential with an R-expression in the discourse. Based on Reinhart (1986), Grodzinsky and Reinhart propose that coreference is ruled out by an LF rule which resolves the competition between syntax and pragmatics:

4. Rule I: Intrasentential Coreference

NP A cannot corefer with NP B if replacing A with C, C a variable A-bound by B, yields an indistinguishable interpretation.

Simply put, Rule I is responsible for excluding coreference between a pronoun and some referring expression whenever the intended interpretation can be achieved through binding. The representation in (3c) is therefore excluded since an A-binding alternative is available (i.e. *Lucie<sub>i</sub> adores herself<sub>i</sub>*).

A requirement of Rule I is that two LF representations be held in working memory in order to be compared for truth-conditional equivalence. This task, according to Grodzinsky and Reinhart, is beyond children's computational capacity and they are therefore forced to opt for a guessing strategy, which results in their performance being at chance level. This claim is backed by psycholinguistic evidence drawn from priming experiments which showed that children, as well as agrammatic aphasics (who also display DPBE) have a limited working memory compared to the adult population (Grodzinsky et al. 1989).

The experiments that claimed to test children's knowledge of Principle B were in fact testing their ability to apply Rule I (see Chien and Wexler 1990 for a similar claim, though due to a different analysis). Children's non-adult-like performance on sentences such as (1b) does not mean that they have not acquired Principle B and it can thus be maintained that principle B is innate.

Further support is provided by experimental results reported in Chien and Wexler (1990) who found that children's performance improves dramatically when tested on sentences where the antecedent is a quantified NP:

5. Every bear touches him

Since quantified NPs are non-referential, they cannot, by definition, corefer with anything. Unlike the example in (2) where the antecedent was referential, we find that there is only one way in which the quantified NP can relate to the pronoun, namely via a binding relation, illustrated here in (6):

6.  $\forall x (Bx \rightarrow x \text{ touches } x)$

Consequently, interpreting (5) does not involve Rule I and therefore children's performance is expected to pattern more closely with that of adults. This is indeed what we find.

## 2.2 Absence of DPBE: Greek

Recent research has shown that, perhaps unexpectedly, Greek seems to lack DPBE altogether (Varlokosta 2000, 2001). Varlokosta (2001) tested 20 Greek children between the ages of 3;7 and 5;9 using the Truth Value Judgement Task (Crain and Thornton 1998). This task is designed to elicit grammaticality judgements from children by presenting them with context-utterance pairs and asking whether the sentence is a 'true' or 'false' description of the context. The context is created in the form of a short puppet-story. At the end of each story one of the puppets describes what has just happened asking the child to judge whether the description was correct. The sentences chosen have two possible interpretations, only one of which is compatible with the adult structure. The child's grammar is taken to diverge from the adult's if s/he accepts the non-adult interpretation. On the other hand, the child is taken to have achieved the adult setting if s/he rejects it.

Below is an example of the sentences used (from Varlokosta 2001:784):

7. O Goofy<sub>i</sub> skepase afton<sub>y/\*i</sub>  
the Goofy covered him  
'Goofy covered him'

Surprisingly, the children performed in a manner consistent with the target grammar responding correctly 87% of the time. These facts have important implications for children who are exposed to Greek as well as to a language that clearly displays DPBE, such as English. Since Greek lacks DPBE, Greek-English bilinguals could be expected to display acceleration effects which would allow them to perform better than English monolinguals in Principle B contexts.

Researchers have investigated this possibility as a way of testing whether transfer effects can occur at all. The assumption underlying such research has often been made clear: if acceleration is attested and the children do master Principle B earlier than the respective monolinguals, then transfer has occurred and the two systems must be able to interfere with each other. On the other hand, if acceleration does not obtain, it is proof of the child's inability to transfer from one language to the other, due to the two systems developing completely separately



(Varlokosta and Dullaart 2001; Sanoudaki 2002). These, however, are the only two logical possibilities so long as we restrict our options to either the Single System Hypothesis or the Separate Development Hypothesis.

On the other hand, the model proposed in chapter 1 maintains that transfer effects are the result of Cascade-updating, a strategy used by the language system to propagate information collected during the acquisition process. Cascade-updating is triggered when new information is associated to some members of a paradigm and it has the effect of distributing such information to all the other members of that same paradigm, unless some conflicting properties interfere. Because this architecture restricts generalisation to other members of a paradigm, it dramatically reduces the cases in which transfer effects can be expected. Although it is a truism that if acceleration should be attested it would confirm interference between the two languages, once we assume Cascade-updating it no longer follows that lack of acceleration demonstrates lack of interference. It may just be that the environment necessary to trigger Cascade-updating never arose.

Since the triggering of Cascade-updating is dependent on paradigmatic properties, it is of fundamental importance that we establish what properties may be involved before any predictions can be made. It is to the properties of Greek pronouns that we now turn.

## 2.3 Greek Pronouns

Some researchers have argued that the lack of DPBE in Greek pronouns may be due to their lexical specification (Baauw et al. 1997; Varlokosta 2000, 2001; Sanoudaki 2002). Following Cardinaletti and Starke (1999b), it has been suggested that, strictly speaking, Greek does not have third person pronouns, the elements commonly referred to as pronouns being in fact demonstratives:

8. (from Sanoudaki 2002;107/8)

- i) Demonstratives always have a special morphological marker, never found on personal pronouns: the strong pronoun *aftos*, *afti*, *afto* is morphologically identical to the demonstrative pronoun *aftos* 'this'
- ii) Demonstratives may refer to non-human entities in contexts requiring strong forms (personal pronouns cannot):

O Yianis vlepi afton. (afton= Bill/ the computer)  
the John sees him-strong pronoun.

'John sees him'

- iii) Demonstratives, contrary to personal pronouns, cannot overrule their disjointness requirement through accidental coreference -  
Greek speakers in general reject the reflexive reading of [the sentence below], with the meaning 'John loves only himself', even in adequate pragmatic context:

*O Yianis agapa mono afton*  
the John loves only ?-3.SG.MASC.ACC

- iv) Demonstratives typically make spatial distinctions of the far/near type, while pronouns seem never to do that - such a distinction in Greek is probably the *aftos-ekinos* one, where *ekinos* is the 'far' type.

The properties listed above suggest that the Greek element *aftos* is in fact a demonstrative. This provides a possible explanation for the near-adult performance of Greek children when tested in contexts such as (7): if DPBE may only occur with pronouns, it follows that a language that lacks third person pronouns will also lack DPBE.

However, there is one context in which the Greek element *aftos* seems to behave just like a personal pronoun. As the example in (9) suggests, *aftos* seems to be able to enter a binding relation with a c-commanding antecedent, a property that is not usually associated with demonstratives.

9.     *Kathe andras nomizi oti aftos ine eksipnos*  
      every man   thinks that he   is   clever  
      'Every man thinks that he is clever'

As demonstratives are generally taken to be R-expressions (see for example Buring 2005, Diesing 1996), we would expect their distribution to be regulated by Principle C of Binding Theory, whereby they must be disjoint from any c-commanding antecedent. Nevertheless, (9) suggests that there is at least one context in which the Greek element *aftos* seems to behave just like personal pronouns in that it appears to be subject to Principle B instead. However, there is some evidence that the sentence in (9) does not actually involve binding of the demonstrative.

Firstly, it must be noted that – being a pro-drop language – Greek allows binding of a phonologically empty pronoun, typically identified as *pro*:

10. [Kathe andras]<sub>i</sub> nomizi oti *pro*<sub>i/k</sub> ine eksipnos  
every man thinks that *pro* is clever  
‘Every man thinks that he is clever’

In fact, the example in (10) represents the unmarked case, while (9) is generally used to convey a contrastive reading, roughly translatable as ‘every man thinks that he is clever’.

This raises the question of whether *pro*-binding could also be responsible for the availability of a bound reading in (9). If this were the case, the demonstrative *aftos* would remain unbound, as required by Principle C and in accordance with its demonstrative status. Its presence in the structure could be independently explained in terms of pragmatics, given that it yields a contrastive reading.

Indeed, it is well-known that Greek demonstratives may co-occur with full DPs, where they occupy the specifier position (from Panagioditis 1999: 718. See also Brugè 2002, Campbell 1996, Giusti 2002):

11. [<sub>SpecDP</sub> aftos [<sub>D</sub> o [<sub>NP</sub> neos andras]]]  
this the young man  
‘This young man’

Applying the structure in (11) to (9) above we obtain (12), whereby *aftos* co-occurs with *pro*, the latter being the element that undergoes binding by the quantified NP:

12. [Kathe andras]<sub>i</sub> nomizi oti aftos *pro*<sub>i/k</sub> ine eksipnos

If we accept this analysis, the sentence in (9) does not present a problem for the demonstrative status of *aftos*.

In sum, the Greek element *aftos* shares a number of distributional properties with demonstratives: it refers to non-human entities in contexts requiring strong forms, it makes a spatial distinction of the far/near type and it strongly resists accidental coreference. Since DPBE may only occur with pronominal elements, it follows that if a language lacks third person pronouns it will also lack DPBE<sup>1</sup>. This

seems to be the case in Greek. In the next section I will look at what the implications are with regard to transfer in bilingual acquisition.

## 2.4 Implications for Transfer

Following the conditions imposed by Cascade-updating, any information the bilingual child has acquired through exposure to the demonstrative *aftos* can be transferred to elements in the other L1 provided that these happen to share the relevant paradigmatic properties. As a result, a question arises with regard to children simultaneously acquiring Greek and a pronoun language:

13. Would bilingual children transfer their knowledge of the Greek paradigm to their other L1 thereby displaying acceleration effects?

Following the discussion in section 2.3, for (13) to be answered positively it is necessary that the pronouns in the other L1 inherit demonstrative status. In order to establish whether this is a possible outcome within the Cascade-updating model, we must first determine what lexical properties these elements bear.

### 2.4.1 Feature Specifications

It is generally assumed that the distinction between demonstratives and personal pronouns “is not parallel to a within-category distinction” (Cardinaletti and Starke 1999b:288. See also Diessel 1999). In fact, recent discussions in the literature suggest that pronouns and demonstratives may realise different syntactic objects altogether. On the one hand, there is general agreement that demonstratives instantiate functional XPs (Brugè 2002, Giusti 1993, 2002, Noguchi 1997, Panagiotidis 2000, among others). This idea that demonstratives are phrasal in nature can be straightforwardly incorporated within a representationally modular system. This is because within such a system there is no commitment as to what type of structure can be associated with a spell-out rule (cf. chapter 1, section 2.2). The syntactic representations that are mapped onto phonological information may involve chunks larger than terminal nodes (see also Neeleman and Szendrői 2005). Consequently, the suggestion that demonstratives are phrasal can be represented as follows:

14.      +referential                  ↔      FP[3 sg, masc]                  ↔      /afton/

Pronouns, on the other hand, have been generally assumed to be D heads (see for example Postal 1969; Abney 1987; Dechaine and Wiltschko 2002). However, this analysis seems to be forced by theory-internal considerations and its empirical basis has been recently brought into question (see for example Neeleman and Szendrői 2005, Weerman and Evers-Vermeul 2002). The proposal that pronouns are heads began as an attempt to reconcile the traditional idea that pronouns stand for complete nominal phrases, with the theory-internal need to associate each phonological word with either a head or a complement position. This is also necessary within traditional lexical models as these restrict phonological mapping to syntactic terminals.

Consequently, it has been assumed that pronouns are determiners that head a DP containing a null NP. The occurrence of expressions such as (15) has often been analysed as evidence in favour of this view (see for example Abney 1987, Dechaine and Wiltschko 2002).

- 15. a. us linguists
- b. I Claudius

On closer inspection, however, these examples display some radically different properties when compared to standard DPs, which suggests that they may be better analysed as instances of apposition rather than complementation. Firstly, the personal pronouns in (15) do not restrict the interpretation of the following noun (Diessel 1999). In a standard DP such as *the boys*, the determiner's task is that of restricting nominal interpretation by picking out a specific subset of boys from the universe of discourse. The relation instantiated in (15), however, represents the mirror image of this, as it is the noun/proper name that provides some restriction by further specifying who the pronoun is meant to refer to. In other words, a noun/proper name is added to the pronoun in order to restrict the group of potential referents. In standard DPs, on the other hand, restriction applies in the opposite direction.

Secondly, claiming that pronouns are determiners raises the question of why singular pronouns cannot combine with a (common) noun:

- 16. a. \*I linguist
- b. \*you linguist
- c. \*him linguist

This has been recognised as an empirical problem by, among others, Dechaine and Wiltschko (2002:421).

Thirdly, the stress pattern we find in the examples in (15) does not match that of standard DPs. For example, a DP in object position often involves stressing the noun with consequent phonological weakening of the determiner, as in (17) (underlying indicates stress).

17. a. Social struggle is of no interest to [ə] linguist  
b. Social struggle is of no interest to [ði] arsonist

On the other hand, acceptability of the sequences in (15) is dependent upon the pronouns carrying some stress and thus appearing in their full phonological form, as shown in (18).

18. a. Social struggle is of no interest to [As] linguists /\*[əs] linguists  
b. [aɪ] Claudius /\*[ə] Claudius will lead you to victory

Of course, pronouns in object position can be, and often are, unstressed:

19. a. Social struggle is of no interest to [əs]  
b. Social struggle is of no interest to [As]

Fourthly, if pronouns were determiners we would expect that they should be unable to combine with a full DP, contrary to fact<sup>2</sup>:

20. a. We the people  
b. You the workers

Fifthly, there is a wide range of languages in which a pronoun and its putative complement display morphological properties that differ from those found in regular DPs. For example, in Basque we find that DPs involve overt case-marking of the determiner only, as in (21) (from Saltarelli 1988:78)<sup>3</sup>.

21. mutil batzu      zutik      daude      tabernan  
boy some.ABS on-foot 3.ABS.PRES.be.PL tavern.LOC  
'Some boys are standing around the tavern'

However, when a pronoun occurs with a noun both elements must bear the same overt case-marker, indicating that they are two separate instantiations of DP (from Saltarelli 1988:210).

22. Guk emakumeok geure eskubideak errespeta  
we.ERG woman.PL.ERG we.EMPH.GEN right.PL respect  
ditzatela eskatzen dugu  
3.ABS.PL.AUX.3.ERG.COMP request 3.ABS.AUX.1.PL  
‘We women request that they have respect for our rights’

Similarly, Bhat (2004) reports that in some varieties of Kannada – a Dravidian language – a pronoun-noun sequence involves both elements inflecting for case, whilst determiner-noun sequences display overt case-marking on the noun only.

Returning to Basque, it must also be noted that whilst in (21) the determiner follows the noun, in accordance with Basque’s head-last nature, the pronoun in (22) precedes it, once again suggesting that a different type of relation is at play. Also, as we saw for English, Basque allows pronoun-noun constructions only with a subset of its pronouns.

In view of these facts, the claim that pronouns are D-heads seems to have no empirical support. This hypothesis rests entirely on theory-internal considerations, as the idea that pronouns spell-out full DPs is inconsistent with the X-bar schema. On the other hand, as we have seen in (14), spell-out of phrasal elements can be straightforwardly formalised within a representationally modular system. This can be extended to pronouns, as in (23).

23. -referential ↔ DP[3 sg, masc, acc] ↔ /him/

By assuming a representationally modular system, we maintain the view that pronouns stand for DPs without having to associate them either with D or N. This appears to be a theoretical advantage, given the empirical issues discussed in this section.

#### 2.4.2 Addressing Transfer

Now that we have established what properties pronouns and demonstratives might bear, we are ready to address the question in (13) above. By comparing (14) and

(23), we find that, although the two elements share some of their phi-features, they differ in two crucial respects: (i) their referential properties and (ii) their categorial status. Of these two, the latter is of major importance. Recall that, for (13) to be answered positively, it is necessary that the pronouns in the other L1 inherit demonstrative status. However, this is not a possible outcome within the Cascade-updating model. Firstly, it requires that cascade updating override an existing property by attributing the categorial specification FP to elements that are already specified as belonging to a different category, namely DP<sup>4</sup>. Crucially, this possibility is incompatible with the informationally monotonic nature of the process (cf. chapter 1, section 4.2).

Moreover, recall that categorial features are the minimal condition in identifying a paradigm. As a result, elements whose category is FP will – by definition – not belong to the same paradigm as elements which are mapped onto a DP structure. Because Cascade-updating can only ever target members of the same paradigm, it follows that its application cannot result in cross-categorial transfer. Consequently, bilingual children are expected to show DPBE in pronoun contexts regardless of whether their other L1 happens to be Greek.

The literature presents two experiments that have been carried out to test this possibility. I will outline each of them in turn.

## **3. DPBE and Bilingualism**

### **3.1 Greek-Dutch Bilinguals**

Varlokosta and Dullaart (2001) tested 10 Greek-Dutch bilingual children aged 3;3 to 7;6 using the Truth Value Judgement Task (see section 2.2). The experiment was subdivided into two parts to allow for testing the children on their knowledge of Greek and Dutch separately. In both cases, the story involved a puppet, Goofy, performing an action on himself (such as covering himself with a blanket). This context was then followed by another puppet uttering a sentence which, in the intended interpretation, would be ungrammatical due to a Principle B violation. The following are examples of the sentences used (adapted from Varlokosta and Dullaart 2001:787):



- |     |   |              |
|-----|---|--------------|
| 24. | O Goofy skepase afton<br>the Goofy covered him                                | <i>Greek</i> |
| 25. | Goofy heeft hem schoongemaakt<br>Goofy has him cleaned<br>'Goofy cleaned him' | <i>Dutch</i> |

26. Greek-Dutch bilinguals (adapted from Varlokosta and Dullaart 2001:787):

The results in (26) converge with those obtained for monolingual Greek and Dutch children respectively (Varlokosta 2000; Philip and Coopmans 1996). This provides a negative answer to (13), which, as we have seen, is not an unexpected outcome under the Cascade-updating model. Because the elements at issue do not belong to the same paradigm, they cannot affect each other and therefore acceleration does not occur.

The same experiment has been replicated with Greek-English bilinguals (Sanoudaki 2002). The subjects were 10 children aged between 4;1 and 6;7 who had been acquiring Greek and English simultaneously from birth. The same experimental technique was used. This experiment was also subdivided into two parts to allow for testing the children on their knowledge of Greek and English separately. In this case the story involved a toy character who would cover himself with a blanket. The following are examples of the sentences used:

27.    a. O    Goofy skepase afton                      *Greek*  
             the Goofy covered him
- b. Grandpa covered him                      *English*

As in the previous experiment, the children performed in a highly adult-like fashion on the test conditions for Greek. However, their performance on the English sentences was just above chance:

28.    Greek-English bilinguals (from Sanoudaki 2002;115)

<i>Language</i>	<i>Lexical Item</i>	<i>Correct responses</i>
Greek	Demonstrative	16/20 (80%)
English	Strong Pronoun	12/20 (60%)

Once again, the results in (28) converge with those obtained for monolingual Greek and English children respectively (Varlokosta 2000, Chien and Wexler 1990), confirming that no-transfer has taken place.

### 3.3 Interlude

Section 3 set out to test a major claim made by the model presented in chapter 1, namely that transfer effects only affect items of the same paradigm. In the literature on BFLA it has sometimes been assumed that transfer ought to be attested whenever the two languages differ with regard to some domain. Based on this assumption, research has been carried out on bilingual children acquiring language pairs which differ with regard to DPBE, such as Greek-Dutch and Greek-English. Because monolingual Greek children do not display DPBE, it was concluded that, if transfer is at all possible, bilingual children would employ this advantage thereby improving their performance on Principle B compared to Dutch and English monolinguals of the same age. Even though it is widely agreed that the Greek elements are underlyingly different from their English and Dutch counterparts, this has been tacitly assumed to be of minor importance compared to similarities on the surface.

In contrast, the model proposed in chapter 1 claims that transfer effects are the result of an updating mechanism which can only affect members of the same paradigm. This gives important theoretical status to the categorial difference that

has been claimed to exist between demonstratives and pronouns. Because the two sets of elements are mapped onto different categorial features they are in fact expected to develop totally independently of each other. The results presented in (26) and (28) provide evidence in favour of this view.

## 4. DPBE in Monolinguals

### 4.1 Absence of DPBE: clitics

#### 4.1.1 Introduction

Researchers have reported a number of experiments aimed at testing whether the acquisition of clitics patterns with that of pronouns with regard to Principle B. Notably, McKee (1992) tested a group of children acquiring a pronoun language (English) and another group acquiring a clitic<sup>5</sup> language (Italian). The experiments were carried out using the Truth Value Judgement Task (see section 2.2).

#### 4.1.2 Experimental Evidence

McKee tested 60 children between the age of 2;6 and 5;3 who were acquiring English and 30 children between the age of 3;7 and 5;5 who were acquiring Italian as their first language. After staging a short story, a puppet would utter a sentence which, in the intended interpretation, is ungrammatical in the adult language. The following are examples of the sentences used (From McKee 1992:36):

- |     |   |                |
|-----|---|----------------|
| 29. | Smurfette washed her  | <i>English</i> |
| 30. | Lo gnomo lo lava<br>the elf him-CL wash.3SG<br>'the elf washes him' | <i>Italian</i> |

English children accepted a co-referential interpretation in sentences such as (29) roughly half of the time, confirming their difficulty with reference resolution (cf. section 2). However, Italian children performed in an adult-like fashion 80% of the time, raising an important question with regard to the acquisition of clitics (see also Escobar and Gavarro 1999, Jakubowicz 1993, Padilla 1990 for similar results in Catalan, French and Spanish, respectively). McKee observes that the lack of DPBE in the acquisition of clitics may be related to the syntactic configuration in which they appear. Although she did not offer a successful explanation for this, the idea

that clitics' syntactic position may be responsible for their lack of DPBE finds some strong empirical support. Before discussing this, however, some clarification on what constitutes a clitic is in order.

#### 4.1.3 Syntactic v. Phonological Clitics

While defining a clitic is no simple matter, it has often been noted, at least since Zwicky (1977) that there seem to be at least two distinct types of cliticisation, namely *syntactic* and *phonological*. Although both types of clitics differ from pronouns in that they are typically unstressed, morphologically reduced, and form a phonological unit with some other element, syntactic clitics also differ in their syntactic behaviour. As is well known, they often appear in complementary distribution with full pronouns. Romance clitics are a typical example of this:

31. a. Gianni vede lui  
       G.     sees him  
       a'. Gianni vede \*lo  
       G.     sees   him-CL
- b. Gianni \*lui vede  
       G.     him sees  
       b'. Gianni lo     vede  
       G.     him-CL sees

This is not the case with phonological clitics:

32. a. John saw him  
       a'. John saw əm  
       b. \*John him saw  
       b'. \*John əm saw

Moreover, phonological clitics pattern with full forms in a predictable manner based on the general phonological processes of the language, while the form of syntactic clitics seems to be entirely a lexical matter. For example, the difference between the full form [ju:] and the phonological clitic [jə] follows the "ordinary reduction of unaccented vowels" (Zwicky, 1977:26) e.g. [tu:] – [tə] for the preposition/marker *to*.

(Compare also the relation between [ðəm] / [əm] / [m̩] with ['dʒenrəl] / [dʒə'nerɪk] and the development seen, for example, in [beɪkən] / [beɪkŋ]).

It is also interesting to note that while phonological clitics can under no circumstances bear any type of stress, syntactic clitics do lend themselves to contrastive stress:

33. a. A: But yesterday you said that you know 'er  
 B: No, I said I know him / \*əm
- b. A: Ieri hai detto che la conosci  
 Yesterday have.2SG said that her-CL know.2SG  
 B: No, ho detto che lo conosco  
 No have.1SG said that him-CL know.1SG  
 'yesterday you said that you know her'  
 'No, I said that I know him'

Although both types of clitics are unstressed in unmarked contexts, it appears that phonological clitics are in an allomorphic relation with full forms, lack of stress being the necessary condition for their distribution. Syntactic clitics, on the other hand, can in particular circumstances bear stress, indicating that lack of stress is probably a consequence of their functional status rather than a defining property of syntactic cliticisation as such.

Even though this overview is far from comprehensive, it raises an important question with regard to DPBE. In particular, whether lack of DPBE is a property specific to syntactic clitics, as the Italian and Spanish experiments might suggest, or whether it carries over to phonological clitics as well. If McKee's observation is correct, and lack of DPBE in Italian is due to the syntactic position of the clitic, we would expect phonological clitics to display DPBE just like strong pronouns since, syntactically, they pattern with strong pronouns.

In order to test this hypothesis, Baauw (2002) set up an experiment involving children acquiring Dutch, a language that has phonological clitics of the type exemplified in (32). Using the Truth Value Judgement Task, he tested 15 children between the age of 4;2 and 5;3 who were acquiring Dutch as their first language. The following is an example of the sentences used (From Baauw 2002:129):

34. Jantje heeft 'm getekend  
Jantje has 'm drawn  
'J. has drawn 'm'

It was reported that the children rejected a co-reference reading only 53% of the time. These results pattern with those reported for the acquisition of Dutch full pronouns (Philip and Coopmans 1996) and thus eliminate the possibility that lack of DPBE could be a property of clitics in general. As a result, the observation that lack of DPBE in Romance clitics may be tied to their syntactic distribution gains some important empirical support. From this point forward, therefore, I will use the term 'clitic' to refer to syntactic clitics (Zwicky's *special clitics*) unless otherwise stated.

#### 4.1.4 Monolingual Transfer

The results reported by Baauw (2002) also raise an important question in relation to potential cases of monolingual transfer (i.e. overgeneralisation). Similarly to the bilingual cases (cf. section 3), we might expect that children acquiring a language that has both clitics and pronouns would generalise across the two sets of elements since these share some core linguistic properties, such as being theta marked as the object of a verb and having their interpretation regulated by Binding Theory. If this generalisation does occur, it would help those children acquiring a clitic language to exit the DPBE stage earlier when compared to children acquiring a pronoun language.

On the other hand, the Cascade-updating model claims that it should be possible for clitics and pronouns to develop totally independently of each other. This may be the case if the two belong to separate paradigms, as was the case for Greek demonstratives on the one hand and Dutch/English pronouns on the other. However, in order to see what predictions the Cascade-updating model makes in relation to the development of clitics and pronouns, we must first consider what lexical properties these elements might carry and which of these properties may be responsible for the clitic's immunity to DPBE.

#### 4.1.5 Clitic Placement

Whether movement or base-generation can be held responsible for the distribution of syntactic clitics has long been a matter of debate. Probably the strongest motivation for movement analyses such as that of Kayne (1975) is the

complementary distribution between the clitic and its associated theta-position, as exemplified in (31).

However, base-generation analyses have often been proposed in view of clitic doubling phenomena in languages such as Spanish and Greek (see for example Borer 1984, Jaeggli 1982, among others).

An insightful alternative has been developed by Sportiche (1993, 1999) who proposed that the two competing views can be reconciled by assuming that clitic structures involve both movement and base-generation. On the one hand, Sportiche analyses clitics as functional  $X^0$  that head their own projection<sup>6</sup> and are base generated above the verb. On the other hand, he argues that the argument position is filled with an XP that must move to the specifier position of the projection headed by the clitic. This gives rise to the following relation (for ease of presentation, only the relevant elements are included):

35. [CLP XP<sup>^</sup> [[CL lo] [ vede XP\*]]]  
       him-CL see.3SG

Although the structure involves movement of the XP from XP\* to XP<sup>^</sup>, the clitic is analysed as a base-generated head. In languages that exhibit clitic-doubling, the movement is assumed to take place at LF, hence the surface order of *clitic*→*double*. In the case of languages that, like Italian, do not allow clitic doubling, the XP is assumed to be a *pro*.

This proposal neatly incorporates the advantages of both movement and base-generation views within a single analysis. On the one hand, clitic-doubling no longer represents a challenge to the Theta Criterion (Chomsky 1981) as the structure at issue includes only one DP, namely XP\* in (35) above. The phi-agreement we find between the clitic and the doubled DP can also be made to follow, given standard Minimalist assumptions with regard to the Spec-Head configuration. On the other hand, the complementarity of distribution between clitics and DP objects in languages like French and Italian is reduced to a requirement that – in those languages – the head of the CLP and its specifier cannot be simultaneously filled. The morphological differences we find between doubling and non-doubling languages provide some independent motivation for this principle (see Sportiche 1993 for details).

With regard to our investigation, Sportiche's proposal offers us some understanding of the lexical properties that are associated with clitics. Firstly, it

suggests that clitics, unlike pronouns, are functional heads. Moreover, it assumes the presence of a chain relation as a necessary part of clitic configurations and, consequently, of clitic interpretation. Notably, no such relation is claimed to exist for phonological clitics (cf. 32). We could therefore tentatively assume that there is a link between lack of DPBE and the chain relation upon which clitic interpretation is based<sup>7</sup>:

36. DPBE is absent in contexts where interpretation of the pronominal element relies on chain formation.

The purpose of the generalisation in (36) is to capture a correlation between lack of DPBE and the clitic's syntactic environment. Interestingly, this finds some empirical support when we look at Serbo-Croatian, a language which has both clitics and obligatorily moved pronouns.

#### 4.1.6 *Serbo-Croatian Pronouns*

According to (36), clitics lack DPBE as a result of the fact that their configuration involves establishing a chain relation. This implies that lack of DPBE should not be peculiar to clitics. If a language requires that full pronouns obligatorily occupy a non-theta position, we expect that these too will lack DPBE, even though they may not be syntactic clitics. This scenario is found in Serbo-Croatian.

Firstly, note that unmarked structures in Serbo-Croatian are SVO if the object is a full DP (Perovic 2004):

37. Marko voli kolace  
'Marko loves cakes'

However, if the object is a pronoun, it must appear preverbally, just like a clitic<sup>8</sup>:

38. a. Marija njega voli  
Mary him loves  
'Mary loves him'  
b. Marija ga voli  
Mary him-CL loves  
'Mary loves him'



Consequently, the interpretation of Serbo-Croatian pronouns must be taken to rely on chain formation, since they do not surface in the post-verbal theta position. If (36) is on the right track, we expect that Serbo-Croatian pronouns – just like clitics – will lack DPBE, since their interpretation relies on chain formation, albeit the configuration they appear in is arguably more straightforward than the one in (35). Experimental evidence suggests that this is indeed the case.

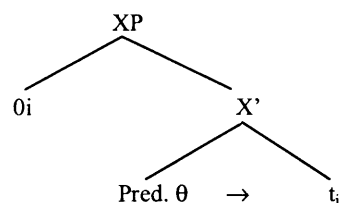
Perovic (2004) tested 37 children aged between 3;3 and 6;11 who were acquiring Serbo-Croatian as their first language and found that they performed in highly adult-like fashion in both clitic (38a) and pronoun (38b) contexts. On the pronoun condition scores varied between 89%-95% while on the clitic condition they were between 85%-95%, depending on age group. These results are in accordance with the generalisation in (36), in that they suggest that lack of DPBE is not peculiar to clitics but correlates with the syntactic structure in which pronominals occur.

#### 4.2 Beyond Generalisation

Ideally, one would like to have an explanation for why (36) should hold. As far as I know, the only attempt at an explanation has been put forth by Baauw (2002) who, building on Sportiche (1993, 1999), offers an account in terms of semantic interpretation.

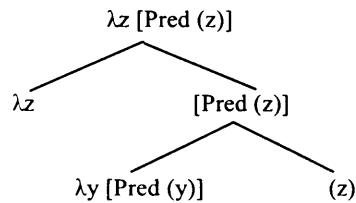
Baauw (2002) suggests that the DP movement analysis proposed by Sportiche be viewed in terms of lambda-abstraction, following Neeleman and Weerman's (1999) analysis of passives. On this view, passive constructions are an example of null-operator movement whereby the predicate's internal theta-role is assigned to an empty operator that subsequently moves to a position above the verb. This structure is represented as in (39).

39.



A further assumption on which this analysis is based is that the head of the chain corresponds to the introduction of a lambda operator at Logical Form (LF). The LF of (39) is therefore assumed to be as in (40):

40.

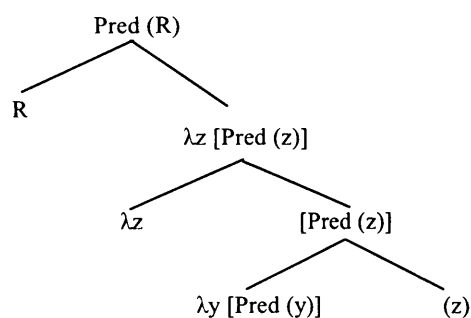


The derivation in (40) can be summarised as follows:

41.
  - a. The predicate and a variable (i.e. the trace) are in a sisterhood relation.
  - b. The value of the variable is therefore assigned to the argument position, and the lambda operator is removed via lambda reduction, yielding [Pred (z)].
  - c. The resulting structure [Pred (z)] is now in a sisterhood relation with a lambda operator (following the assumption that null-operator movement introduces a lambda operator at LF).

The environment is now right for more reduction to take place, under sisterhood with some R-expression, as in (42).

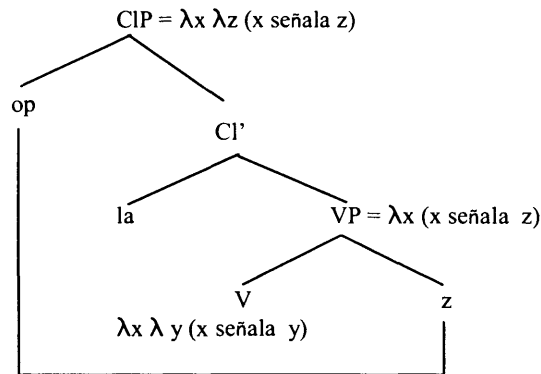
42.



The result is that R will be interpreted as the internal argument of Pred despite it not having originated in that position. This is because the process of lambda-reduction involves associating the R-expression with a variable which had previously been associated with the trace of the operator chain (cf. 41b).

Baauw (2002) proposes that the same procedure could be applied to clitic constructions, as in (43) (adapted from Baauw 2002;133).

43.    *la*    *niña*  $Op_i$  *la*        *señala*     $t_i$   
          the girl        her-CL points-at

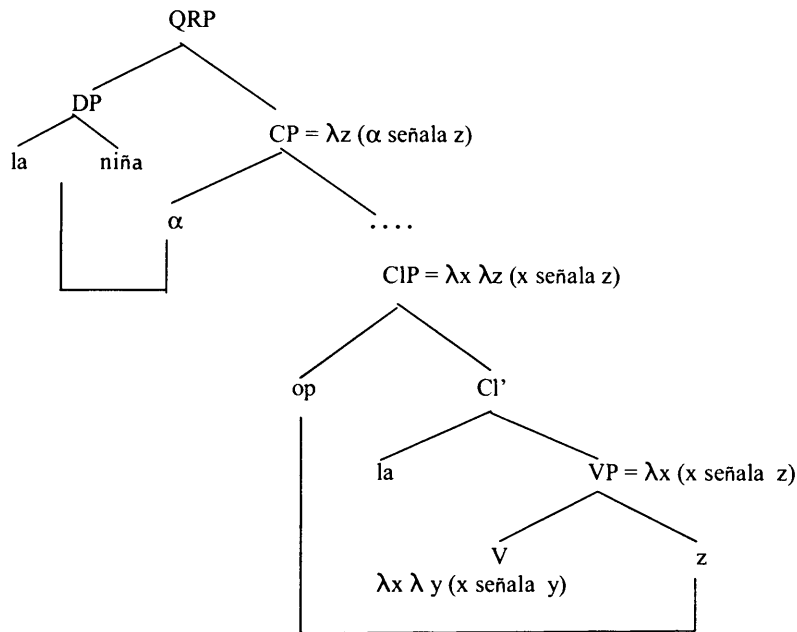


Note that the operator structure is equivalent to the construction proposed by Sportiche (1993, 1999) (cf. section 4.1.5). The relevant derivation for (42) is given in (44).

44.    a. At the VP level, the variable *y* is associated with the operator trace,  
          indicated here by the variable *z*, with consequent removal of the  
          lambda operator via lambda reduction (cf. 41b).  
        b. At CIP level, the moved null-operator introduces a lambda operator,  
          (cf. 41c).

When the resulting structure is applied to the external argument *la niña*, a Principle B violation follows, provided that we also assume that the DP *la niña* undergoes QR. The complete derivation is given in (45) (only those layers that are immediately relevant to the discussion are included in the tree structure).

45.



At the CP level, the trace left behind by the Quantifier Raised DP – indicated in (45) by the variable  $\alpha$  – provides a value that can be assigned to the external argument of the predicate. This triggers further lambda reduction, thus yielding the structure  $\lambda z (\alpha \text{ señala } z)$ . At this point, Baauw suggests that the two variables  $\alpha$  and  $z$ , representing the trace of the DP and the trace of the null-operator respectively, are identified as the same ( $\alpha = z$ ) since, by assumption, they share the same features. This yields the structure  $\lambda z (z \text{ señala } z)$ , with consequent Principle B violation.

According to the derivation in (45), the only way in which a clitic can have the same referent as the subject (i.e. *la niña*) is via binding. As this does not involve application of Rule I, DPBE is predicted to be absent in clitic contexts, the desired result.

It must be noted that this account is not without problems. A claim implicit in the structure in (45) is that clitics cannot be interpreted as free variables. However, clitics may of course corefer with a DP that appeared earlier in the discourse:

46. Gianni<sub>i</sub> lo<sub>k</sub> conoce  
John him-CL know.3SG

The only possible binder available in (46) seems to be the DP *Gianni*. Nevertheless, *Gianni* cannot bind the clitic as this would give rise to a Principle B violation. This

does not make (46) ungrammatical, however, as the clitic could be referring to some discourse-given entity other than *Gianni*. In order to deal with this problem, Baauw suggests that an empty topic operator could be assumed, thus introducing an inter-sentential, non-local binder for the clitic.

47. [TOPIC ec]<sub>i</sub> Gianni lo<sub>i</sub> conosce

Although this solution may not be entirely satisfactory, Baauw's analysis does provide us with a potential explanation for the generalisation in (36). Moreover, it does so in a way that enables us to understand why clitic contexts pattern with quantified-NP contexts (cf. section 2.1); DPBE is ruled out in both cases because both cases necessarily involve a bound-variable interpretation.

In sum, clitic constructions involve NP movement to the specifier of the Clitic Phrase. The resulting structure must necessarily be assigned a bound variable interpretation, a process which does not involve application of Rule I. Since, as discussed in section 2.1, DPBE arises from computational limitations related to Rule I application, clitic constructions are correctly predicted to lack DPBE.

## 5. Implications for Overgeneralization

### 5.1 The Role of Lexical Specification

In section 4 we saw that lack of DPBE in clitic contexts is closely linked to the movement relation involved in the interpretation of clitic constructions. Due to their  $F^0$  status, clitics are base-generated in pre-verbal position. By hypothesis, this configuration triggers movement of the object DP from complement of VP to spec-CLP position (cf. 35 above). Lambda-abstraction is then applied at LF (cf. 43), forcing a bound variable interpretation, with consequent immunity to DPBE. Therefore, it seems that the lexical property that is ultimately responsible for clitics' immunity to DPBE is their status as  $F^0$ .

We are now ready to address the issue presented in section 4.1.4 with regard to the potential occurrence of overgeneralisation in languages that have both clitics and pronouns. The question that arises is reminiscent of the one we asked earlier with regard to bilinguals (cf. 13):

48. Can immunity to DPBE be transferred to pronouns (and thus cause acceleration effects)?

Let us consider what lexical properties are associated with the elements at issue. Following the discussion in section 4, the lexical mapping for the Italian clitic *lo* would be as in (49).

49.      -referential                  ↔      F[3 sg, masc, acc]                  ↔      /lo/

This differs from the mapping associated with the pronoun *lui* (cf. section 2.4.1), as represented in (50).

50.      -referential                  ↔      DP[3 sg, masc, acc]                  ↔      /lui/

Importantly, the two elements differ in relation to their categorial specification, in a manner which is reminiscent of, though radically different from, the pronoun/demonstrative distinction discussed in section 2.4.1. Consequently, the Cascade-updating model makes the same prediction for the monolingual cases as it did for the bilingual cases (cf. section 2.4.2); whatever knowledge the child might have gained with regard to clitics cannot be transferred to pronouns, as this would involve cross-categorial transfer, an impossible outcome within this model. The answer to (48) is therefore expected to be negative. Consequently, children are expected to show DPBE in pronoun contexts regardless of whether the language they are acquiring also happens to have clitics.

In the next section I will present an experiment conducted with children acquiring Italian as their first language which was set up to test this prediction.

## 5.2 The Experiment

The goal of the experiment was to find out whether Italian speaking children show DPBE in pronoun contexts (51a) and clitic contexts (51b):

51.    a. La    mamma sta asciugando lei                                  (NPX-N)  
          the mom    is drying           her
- b. La    mamma la           sta asciugando                                  (NCX-N)  
          the mom    her-CL is drying  
          'The mom is drying her'

The NPX-N condition was meant to test children's ability to interpret the pronoun as disjoint from the preceding DP. In the same vein, the NCX-N condition was meant to test their ability to interpret the clitic as disjoint from the preceding DP.

#### 5.2.1 *The Subjects*

21 children and 3 adults participated in the study. Three more children were excluded as they either did not pay attention or answered 'yes' to most of the control questions. The child subjects were from the Scuola Materna Adele Palli in Voghera (Lombardy region) and they ranged in age between 3;4 and 5;9 (mean 4;8). The children were all brought up in Italy by Italian speaking parents and were acquiring Italian as their first language. The adults were native Italian speakers and were selected from the same region as the children in order to minimise the effects of dialectal variation between subjects.

#### 5.2.2 *Material and Procedure*

The experimental technique used was the Yes/No Judgment Task adapted from Chien and Wexler (1990). This is a comprehension test used to elicit grammaticality judgements from children. The task is used to test whether, for a given structure, children allow interpretations that are ruled out by adults.

The experiment consisted of 2 test conditions: one for clitics (NCX-N) and one for pronouns (NPX-N). For each experimental condition there were 6 different trials with syntactically identical inputs but which involved 3 different sets of characters and 3 different actions. The three predicates used were: *asciugare* (dry) *lavare* (wash) *pungere* (sting/poke) each of which appeared twice for the test conditions. After presenting the child with a cartoon picture, the experimenter gives a context-setting input by pointing out the characters to the child. Finally, the experimenter asks a question related to the picture to which the child is expected to give a yes/no answer. For example, the experimenter would present the child with a picture of two male characters (clearly representing an adult and a child) one of whom was performing some action. An example of the trial is given in (52):

52. Test condition for pronoun context (NPX-N)



*Experimenter:* "Questo e' Pierino seduto per terra e questo e' il vigile dentro la vasca da bagno. Il vigile sta lavando lui?"

(This is Little Peter sitting on the floor and this is a policeman in a bathtub. "Is the policeman washing him?")

*Adult response:* No

The question in (52) leaves the child with a choice between two meaning-utterance pairs: one is an accurate description of what happened but it is ruled out by Principle B. The other is a grammatical but inaccurate description of the story:

53. Meaning<sub>1</sub>: the policeman<sub>i</sub> is washing him<sub>i</sub> [True, though ruled out by Principle B]

Meaning<sub>2</sub>: the policeman<sub>i</sub> is washing him<sub>k</sub> [Grammatical but False]

The child's answer is taken to represent his/her competence: children answering "yes" are taken to allow meaning<sub>1</sub> and therefore to diverge from adult competence as far as Principle B is concerned.

### 5.2.3 Control Conditions

The sentence-meaning pairs described above include a true-ungrammatical and a false-grammatical option. In addition to these, three control conditions were also included. These consisted of two sets of true-grammatical pairs, one for clitics (NCM-Y) and one for pronouns (NPM-Y), as well as one set of false-ungrammatical pairs (ConX-N). The 30 items were arranged in a pseudo-random order<sup>9</sup>.



#### 5.2.4 True-grammatical

A “yes” response to (52) is reconcilable with an adult-grammar. It might just be the case that, although the child considers the sentence ungrammatical, his/her judgement is overridden by the truth value of the sentence. In other words, s/he understands that the speaker must have meant “the policeman is washing himself” and therefore answers “yes”. In order to control for this, utterance-meaning pairs of the type true-grammatical were also included for both pronoun (NPM-Y) and clitic contexts (NCM-Y). If it is the case that truth overrides grammaticality, then the child will treat true-ungrammatical and true-grammatical pairs similarly. However, if the child’s grammar allows two interpretations, s/he will perform at ceiling in the true-grammatical cases but at chance-level in the true-ungrammatical cases.

#### 5.2.5 False-ungrammatical

These sentence-meaning pairs were included randomly to ensure that the child’s “yes” responses were not due to independent reasons (it is well documented that children have a tendency to answer “yes”, presumably to please the adult). Given that a false-ungrammatical pair must necessarily elicit a “no” answer, they contributed information about the validity of the data. These sentences were coded as CoNX.

#### 5.2.6 Results

The adult subjects performed at ceiling in the control as well as in the test conditions, as given in (54) and (55) respectively:

54. Control conditions: adult subjects.

<b>ConX-N</b>	<b>NPM-Y</b>	<b>NCM-Y</b>
18/18 (100%)	18/18 (100%)	18/18 (100%)

55. Test conditions: adult subjects.

<b>NCX-N</b>	<b>NPX-N</b>
18/18 (100%)	17/18 (94%)

The child subjects too achieved high scores in the control conditions. All children performed in virtually adult-like fashion in the true-grammatical (NPM-Y / NCM-Y) as well as the false-ungrammatical (ConX-N) condition:

## 56. Control conditions: child subjects

ConX-N	NPM-Y	NCM-Y
121/126 (96%)	119/126 (94%)	121/126 (96%)

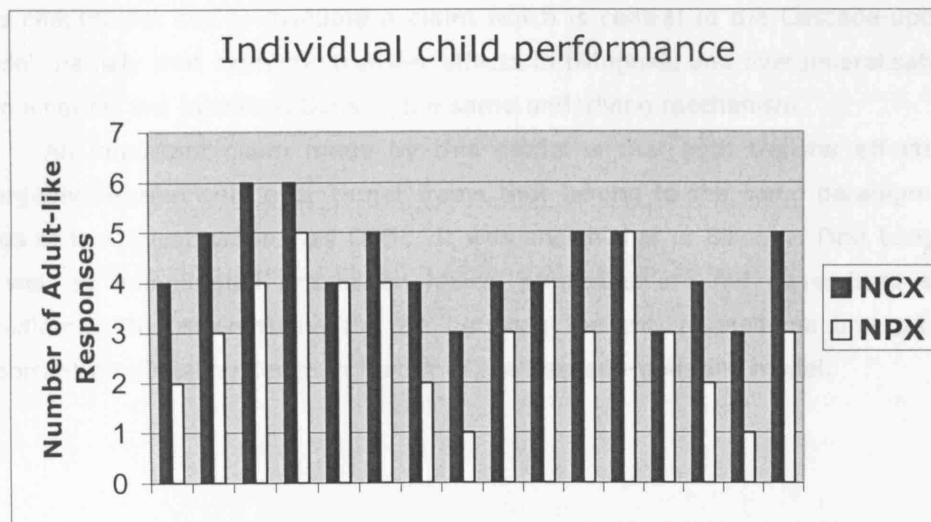
On the test conditions, however, only 5 children performed in adult-like fashion in both pronoun and clitic contexts<sup>10</sup>. Those children who did not perform in adult-like fashion scored quite high in clitic contexts but performed at chance level in pronoun contexts:

## 57. Test conditions: child subjects

NCX-N	NPX-N	NCM-Y	NPM-Y
70/96 (73%)	49/96 (51%)	92/96 (96%)	89/96 (92%)

A t-test revealed that performance in NCX-N and NPX-N contexts differs significantly ( $p=0.001$ ). The chart in (58) shows that the two contexts differed unidirectionally, with none of the children performing higher in NPX-N contexts:

## 58. Individual child performance



At this point, a clarification is in order. It is well known that the use of pronouns in Italian, as well as in other Romance languages, is marked and that a sentence like (51a) requires a particular intonation in order to be acceptable. Although care was taken to ensure that the correct intonation was used, this could - in principle - raise doubts about the validity of the experiment. For example, it could be argued that

the children performed at chance on the NPX-N condition because they simply did not like pronoun use in that context and that this experiment has nothing to do with children's potential difficulty with reference resolution. Nevertheless, this conclusion is challenged by the results in (56). Children's performance on the control condition for pronouns (NPM-Y) indicates that they do not have any problem accepting full pronouns in object position, as they did so 94% of the time when the context did not force a Principle B violation.

Consequently, the experiment presented in this section indicates that Italian children reject Principle B violations significantly more often in clitic contexts than in pronoun contexts. Moreover, on the test condition for pronouns Italian children performed in a manner consistent with children acquiring a language that does not have clitics (e.g. English, cf. section 2). This provides a negative answer to the question in (48) and thus lends further support to the Cascade-updating model by confirming our expectation that lack of DPBE cannot be transferred from clitics to pronouns.

## 6. Conclusions

This chapter set out to evaluate a claim which is central to the Cascade-updating model, namely that syntactic transfer effects in bilinguals and overgeneralisation in monolinguals are manifestations of the same underlying mechanism.

An important claim made by this model is that both transfer effects and overgeneralisation only ever target items that belong to the same paradigm. The focus of the investigation was DPBE. It was shown that in Bilingual First Language as well as monolingual acquisition lexical properties are not generalised across paradigms. This strengthens the tie between the two phenomena and provides important empirical evidence in favour of the Cascade-updating model.

## NOTES

---

<sup>1</sup> Two clarifications are in order. Firstly, Greek cannot be said to lack third person pronouns altogether, as its pronominal inventory arguably includes *pro*. However, this is tangential to our discussion since the experiments at issue involve accusative contexts and are thus independently incompatible with *pro*, which is typically allowed only in the nominative.

Secondly, the claim that lack of DPBE in Greek is due to the lack of third person pronouns predicts that DPBE may occur with first and second person pronouns. However, it is generally agreed that first and second person contexts are not testable due to the fact that, unlike third person contexts, they only allow one interpretation in the target grammar (see Crain and Thornton 1998 for a detailed discussion). Consequently, the literature does not report any data on children's handling of Principle B in first and second person contexts.

<sup>2</sup> I am indebted to Ad Neeleman for pointing out these examples to me.

<sup>3</sup> This is based on the assumption that quantifiers are D-heads, since Basque does not have definite/indefinite articles.

<sup>4</sup> Note that the same would be true even if we were to agree that pronouns are D-heads, since this would still separate them from elements that belong to the FP category.

<sup>5</sup> The term clitic is intended as shorthand for "pronominal object clitic".

<sup>6</sup> This view has been echoed by a number of researchers and it is now common practice to categorise clitics as heads. See also Belletti 1999, Franks and Rudin 2004, Kayne 1991, Papangeli 2000, Uriagereka 1995, among many others.

<sup>7</sup> Please note that this generalisation abstracts away from instances of A'-type relations.

<sup>8</sup> Note that the SVO order is allowed in cases where the pronoun is used deictically, while clitics are never allowed postverbally, presumably because they are in complementary distribution with elements carrying new information.

<sup>9</sup> Though the order of the items was first randomised, some of the resulting sequences involved the same (adult) response for three or more items in a row. In order to avoid potential repetition effects, these items were broken up by inserting an item that elicited a different (adult) response.

<sup>10</sup> Although three of these children were part of the older group (5;6 – 5;9) there was no clear correlation between age and performance as some of the children who did not perform in adult-like fashion were also approaching their sixth birthday. Presumably, the different performances were linked to their linguistic rather than chronological age, though this could not be established with certainty as they had past the stage where MLU could be measured and I had no access to the history of their linguistic development.

## CHAPTER 4

---

# On the Immunity of Pro-Drop and Root-Infinitives

### 1. Introduction

In the previous chapter some evidence was given in defence of the idea that the emergence of transfer effects is dependent on the sharing of certain lexical properties. It was argued that, in accordance with Paradigm Formation, paradigmatic identity is a necessary condition for transfer effects to occur across a given group of items and that transfer effects are impossible unless this condition is met.

As typological differences do not necessarily provide grounds for transfer effects, the absence of transfer within a particular domain tells us nothing about the putative independence of the two languages (*contra* claims by De Houwer 1990, Paradis Genesee 1996, Varlokosta and Dullaart 2001, among many others). In the same vein, similarities between structures in the two languages are not a sufficient condition for the occurrence of transfer effects. As we have seen in chapter 3, two linguistic objects may have crucial representational differences despite their displaying a number of similarities on the surface. Because the Cascade-updating model relies on some specific representational similarities in order for transfer effects to arise, the focus of inquiry must move away from concerns about linguistic behaviour and towards issues of linguistic representation. In other words, BFLA becomes heavily dependent on linguistic theorising.

In this chapter I will argue that due to the developments that the Principles and Parameters framework has undergone in the last decade, some typological differences are predicted to be immune to transfer effects once the Cascade-updating model is assumed. So far we have only considered the possibility that transfer effects may fail to arise due to lack of paradigmatic identity (clause i of Cascade-updating) or to lexical incompatibility (clause ii). However, there is at least

one other implication to consider when searching for potential transfer domains. Because Cascade-updating only targets lexical properties, any typological characteristic which is not lexically encoded must necessarily be unaffected by its application. This involves, for example, surface properties that arise from the interaction of linguistic principles which are not lexically encoded. The aim of this chapter is to introduce two such properties and to review some empirical evidence regarding the status of transfer effects within these domains. The chapter is organised as follows.

In section 2 I will discuss the pro-drop phenomenon. I will argue that pro-drop results from a property that is not directly encoded in the lexicon, namely rich agreement, and that it is therefore expected to be immune to transfer. I will then discuss experimental evidence which suggests that this is indeed the case. I will also attempt to shed some light on the notoriously elusive notion of 'rich agreement' by proposing a definition based on the realisation of formal specifications, thus making some original contribution to the theory of pro-drop.

Section 3 looks at the phenomenon known as Root Infinitives. Following work by Hyams (1996) and Wexler (1998), it is assumed that the occurrence of Root Infinitives in child language is linked to properties of the agreement system. This, in turn, leads to the conclusion that Root Infinitives cannot be affected by Cascade-updating. Experimental evidence is discussed which shows that this seems to be the case. The same section also includes a comparison between the Cascade-updating model and Hulk and Müller's (2000) interface hypothesis. It is argued that, within the context of Root Infinitives, Hulk and Müller's (2000) proposal does not score as well as the Cascade-updating model.

Finally, section 4 gives a summary of the overall conclusions.

### **1.1 More on the lack of transfer**

Recall that the Cascade-updating model is taken to be part of a linguistic system that is Representationally Modular in the sense of Jackendoff (1997). As outlined in chapter 1, within this system the focus of the acquisition process is on two separate – yet connected – tasks (repeated from chapter 1):

1.   a.   Acquiring combinations of properties relevant to each separate component.
- b.   Mapping those combinations of properties onto other components.

It is during these processes that Cascade-updating is assumed to operate as a tool for spreading newly acquired knowledge in order to maximise what has been acquired via experience. However, this cannot be the full picture. As already pointed out in chapter 1 (section 3.2.2), correspondence rules do not allow just any type of mapping as they are subject to universal constraints which limit the possible mapping relations that can be established (cf. example 8, chapter 1). Moreover, the idea that crosslinguistic variation is confined to the lexicon does not entail that all typological differences are directly encoded in the lexicon. Some typological differences may emerge from the interaction between language-specific (i.e. lexical) idiosyncrasies and universal principles. In particular, it is now widely assumed that phenomena which were once believed to result from 'parametric' variations are in fact reducible to this type of interaction. One such case is that of pro-drop, which I will address in the following sections.

## 2. The Case of Pro-Drop

A typological difference that has received much attention in the literature concerns the distinction between those languages that allow phonologically null pronominal subjects in finite clauses and those that do not (Chomsky 1981, 1982; Jaeggli and Safir 1989; Rizzi 1982, 1986). A comparison between English and Italian illustrates this difference:

2.     a. \*(He) knows Mary
- b. (lui) conosce Maria

This effect was at first attributed to the idea that the syntax of English and Italian differ with regard to their setting for the pro-drop<sup>1</sup> parameter: English is set to [-pro-drop] while Italian is set to [+pro-drop]. However, as this version of the Principles and Parameters framework was abandoned, subsequent work (following an idea from Borer 1986) has taken the view that syntactic systems do not vary across languages and that crosslinguistic variation is confined to the lexicon. Nevertheless, typological differences such as pro-drop cannot easily be formulated in terms of lexical specification of the items at issue. Even if we assume that the Italian lexicon contains the silent pronominal *pro* while English does not, we still need to appeal to other – more general – principles of the grammar if we are to explain why *pro* cannot also be (successfully) added to the English lexicon.



A number of proposals have been developed to address this issue, all of which take pro-drop to fall out from some general principle which allows the omission of a pronoun in the presence of Person and Number agreement on the verb (see for example Speas' 2006 "Principle of Economy of Projection", Holmberg's 2003 "recoverability of deletion" Rizzi's 1982, 1986 "pro licensing" Roberts' 2003 "deletion under identity"). Every one of these approaches takes pro-drop to be a consequence of some universal rule which – in some form or other – states that pronouns can be left unpronounced (or that *pro* can be inserted, depending on theoretical commitments) in cases where the relevant Person and Number features are already overtly realised somewhere else in the structure. In other words, pro-drop is not encoded in the grammar in any parametric form (i.e. there is no pro-drop parameter as such). Whether or not a language allows pro-drop is a consequence of some universal principle.

This idea is far from new. A number of early generative analyses tried to formalise the intuition that pro-drop is a by-product of "rich" agreement systems (see for example Perlmutter 1971; Taraldsen 1978). This condition can be loosely stated as follows:

### 3. Rich agreement licenses null subjects

Given (3), the distinction exemplified in (2) can be made to follow: Italian will allow pro-drop while English will not, since the former – but not the latter – carries rich agreement on the tensed verb.

Naturally, however, this type of condition immediately raises the question of what qualifies as "rich" agreement. This notion has proved notoriously difficult to pin down, to the extent that many researchers simply assume that (some version of) the principle in (3) holds without further discussion (recently, this was the position taken by Alexiadou and Anagnostopoulou 1998). In the next section I will try to address the issue of richness in some detail.

## 2.1 Remarks on Richness

### 2.1.1 Introduction

Various attempts have been made at defining richness, some of which involve rather elaborate formulations. One example is the proposal put forward by Rohrbacher (1999), according to whom a language can have pro-drop if "in at least one number of one tense of the regular verb paradigms, the person features [1] and [2] are both distinctively marked" (1999:116). Though this seems to make the

correct distinction between – for example – English and Italian, it does not straightforwardly exclude languages like German which also seem to distinguish between first and second person:

4.

	English: poor agreement	Italian: rich agreement	German: ?
1st sg	walk	cammino	spaziere
2nd sg	walk	cammini	spazierst
3rd sg	walks	cammina	spaziert
1st pl	walk	camminiamo	spazieren
2nd pl	walk	camminate	spaziert
3rd pl	walk	camminano	spazieren

Nevertheless, German does not allow pro-drop<sup>2</sup>, as shown in (5):

5. Er sagte dass \*(er) den Hund gesehen hat  
 he said that (he) the dog seen has  
 'he said that he has seen the dog'

This has led a number of researchers to conclude that rich agreement is not a reliable predictor for pro-drop and that the explanation must lie elsewhere (e.g. Jaeggli and Safir 1989, Alexiadou and Anagnostopoulou 1998, among others). However, I would like to argue that it would be too hasty to abandon the idea that – at least within the type of pro-drop I am concerned with – rich agreement plays an important part. Moreover, it must be noted that even if we were to agree that what appears to mark first person in German is in fact a bare stem (certainly a possible assumption and one which is in line with Rohrbacher's generalisation) it would still be desirable to gain some understanding as to why first and second person enjoy such special status. Indeed, I believe that there is some insight to be gained if we are willing to make the two assumptions outlined below.

### 2.1.2 Feature Binarity

With regard to the nature of morphological paradigms<sup>3</sup>, I will assume – following a long tradition – that Person as well as Number features are binary (for more recent formulations see Noyer 1992, and Koenenman 2000, among others). Because the postulation of a Person feature does not allow for a binary distinction, the different

person specifications will be taken to arise from the combination of *speaker* and *addressee* features instead.

The representation of a morphological paradigm which distinguishes between all six forms is as in (6):

6. Italian: Inf: parl-**are** ('to speak')

1sg	parl- <b>o</b>	+sp aad +sg	1pl	parl- <b>iamo</b>	+sp aad -sg
2sg	parl- <b>i</b>	-sp +ad +sg	2pl	parl- <b>ate</b>	-sp +ad -sg
3sg	parl- <b>a</b>	-sp -ad +sg	3pl	parl- <b>ano</b>	-sp -ad -sg

First person singular and plural are underspecified for the feature *addressee*. As Koenenman (2000) points out, this is reflected in the fact that first person plural is ambiguous as to whether or not it includes the *addressee* within its referent<sup>4</sup>. As far as first person singular is concerned, the inclusive reading is automatically ruled out by the [+sg] specification.

### 2.1.3 Richness as Opposition

The idea that rich agreement licenses pro-drop is based on the intuition that the properties of the missing pronoun can be identified through the morphological inflection. In some sense, a morphological paradigm is rich if it can unambiguously instantiate its formal properties. Such instantiation has the consequence that no further specification will be required and therefore no pronoun is needed<sup>5</sup>.

I propose that a morphological paradigm unambiguously instantiates its formal properties if and only if it realises each possible feature opposition. In other words, unambiguous instantiation results if each of the three possible feature types appears in both a positive and a negative setting within the morphological paradigm. If, on the other hand, a paradigm realises only one value plus an underspecified value for some of the features, the system allows for ambiguity. This follows if we consider that an of (underspecified) value can only be distinguished from a specified one, say +f, if alpha is able to acquire the opposite value (minus in this case). However, alpha can only acquire the minus value if some other element in the structure can supply it. Therefore, lack of opposition means obligatory presence of this 'some other element', namely an overt pronoun. The key claim I am making is that the obligatory presence of pronouns in a language reflects a failure of unambiguous instantiation in the verbal paradigm as a whole rather than in individual verbal forms.

Given this assumption, the German paradigm can be distinguished from the Italian one. The former, but not the latter, lacks opposition within the *addressee* as well as the *singular* feature. This is because it does not mark the 1st and 3rd person plural at all, while in 3rd singular and 2nd plural it only marks *speaker*:

7. German paradigm; Inf.: spazier-**en** ('to walk')

1sg	spazier- <b>e</b>	+sp aad +sg	1pl	spazier- <b>en</b>	
2sg	spazier- <b>st</b>	-sp +ad +sg	2pl	spazier- <b>t</b>	-sp aad βsg
3sg	spazier- <b>t</b>	-sp aad βsg	3pl	spazier- <b>en</b>	

As can be seen from the paradigm in (7), German specifies neither the *addressee* nor the *singular* feature in the 3rd person singular: the morpheme *-t* must be underspecified as it is compatible with both 3sg and 2pl (recall that – following Blevins (1995) – syncretism is a consequence of underspecification rather than disjunction of features, see chapter 1 section 2.2).

In fact, German lacks a plural form altogether; for 2nd person plural a general *-speaker* form is inserted. The 1st and 3rd person plural, on the other hand, are realised through a default form which lacks all person and number features, namely the infinitive. Note that the infinitive morpheme cannot be assumed to be simply underspecified for the three features, since it seems counterintuitive to associate person and number features with infinitivals. Also, it seems reasonable to assume that the infinitive too carries some features. For example, Giorgi and Pianesi (1997) argue that what sets apart the infinitive from finite forms is their different value for the feature *perfective*, while Hoekstra and Hyams (1998) suggest a difference with regard to the feature *realised*. However, the issue of what other differences the two groups may have is of no immediate importance to our discussion and will therefore be left unexplored. What does seem relevant is that – unlike finite verbs – infinitives lack person/number features altogether.

A situation similar to the one we saw for German is also found in Icelandic. Assuming that the 1st person singular is not unmarked<sup>6</sup>, the Icelandic paradigm looks as follows:

8. Icelandic paradigm; Inf: seg-**ja** ('to say')

1sg	seg- <b>i</b>	+sp aad +sg	1pl	seg- <b>jum</b>	+sp aad -sg
2sg	seg- <b>ir</b>	-sp aad +sg	2pl	seg- <b>ið</b>	-sp +ad -sg
3sg	seg- <b>ir</b>	-sp aad +sg	3pl	seg- <b>ja</b>	

Although the Icelandic paradigm is slightly richer than the German one (unlike German, Icelandic realises opposition for the *singular* feature) it does not realise opposition for the *addressee* feature since the only person that provides a value for this feature is the 2nd plural. As to the singular, the morpheme *-ir* must be underspecified for *addressee* since it is compatible with both 2nd and 3rd person. Similar to German, Icelandic lacks a form for the 3rd person plural, thus realising a default form in that position.

#### 2.1.4 How Poor is Rich?

It is important to note that the definition of richness I am proposing does not imply that all forms must be distinguishable or that a paradigm must not contain any underspecified forms (*contra*, for example, Rizzi 1982, Koenenman and Neeleman 2001: footnote 23). This point can be illustrated with Portuguese. This language does not distinguish between 2nd and 3rd person plural in a manner reminiscent of the Icelandic system, which – as we saw – does not specify 2nd and 3rd person in the singular (though Icelandic also displays further underspecification, cf. 8). Nevertheless, the Portuguese paradigm still qualifies as rich for it realises full opposition, as can be seen from (9) below:

#### 9. Portuguese paradigm; Inf: compr-**ar** ('to buy')

1sg	compr- <b>o</b>	+sp aad +sg	1pl	compr- <b>amos</b>	+sp aad -sg
2sg	compr- <b>as</b>	-sp +ad +sg	2pl	compr- <b>am</b>	-sp aad -sg
3sg	compr- <b>a</b>	-sp -ad +sg	3pl	compr- <b>am</b>	-sp aad -sg

The feature *speaker* appears in both a plus and a minus setting, as the language distinguishes between, for example, 1st and 2nd person singular. Furthermore, Portuguese also marks the distinction between 2nd and 3rd person singular, thus specifying both values for the *addressee* feature as well.

Like most Romance languages, Portuguese allows pro-drop (From Speas 2006:26)

10. Vi            seu pai    quando passei  
       saw.1SG your father when    passed.1SG  
       'I saw your father when I passed by'

As has been often pointed out, however, pro-drop is not a feature of the Brazilian variety of Portuguese<sup>7</sup> (Rohrbacher 1994:262):

11. \*(Eu) vi        seu   pai   quando \*(eu) passei.  
       I    saw.1SG your father when        I    passed.1SG  
       'I saw your father when I passed by'

Given our formulation of richness, this is in fact expected since the Brazilian Portuguese paradigm does not realise full feature opposition, unlike the European counterpart (cf. 9):

12. Brazilian Portuguese paradigm; Inf.: fal-**ar** ('to speak')
- |     |               |         |     |                  |         |
|-----|---------------|---------|-----|------------------|---------|
| 1sg | fal- <b>o</b> | +sp +sg | 1pl | fal- <b>amos</b> | +sp -sg |
| 2sg | fal- <b>a</b> | -sp +sg | 2pl | fal- <b>am</b>   | -sp -sg |
| 3sg | fal- <b>a</b> | -sp +sg | 3pl | fal- <b>am</b>   | -sp -sg |

As shown in (12), Brazilian Portuguese only spells out two of the three possible features, namely *speaker* and *singular*. This is because this language displays full syncretism between 2nd and 3rd person in both the singular and the plural. As these are the only persons which can potentially motivate presence of the *addressee* feature (if we exclude inclusive 1st person plural, see footnote 4), Brazilian Portuguese fails to spell out the *addressee* feature altogether. Consequently, contrary to European Portuguese, Brazilian Portuguese is not a pro-drop language.

The same point is further illustrated by Old English. In this language, the plural forms were all underspecified for the *speaker* as well as the *addressee* feature, as shown in (13) (from Campbell 1959):

13. Old English paradigm: Inf.: help-**an** ('to help')
- |     |                 |             |     |                 |             |
|-----|-----------------|-------------|-----|-----------------|-------------|
| 1sg | help- <b>e</b>  | +sp aad +sg | 1pl | help- <b>eþ</b> | asp βad -sg |
| 2sg | hilp- <b>st</b> | -sp +ad +sg | 2pl | help- <b>eþ</b> | asp βad -sg |
| 3sg | hilp- <b>þ</b>  | -sp -ad +sg | 3pl | help- <b>eþ</b> | asp βad -sg |

*Prima facie*, the Old English paradigm may look poorer than the Icelandic one (cf.8), as it employs only one suffix in the plural. However, when looking at its feature specification we can see that Old English did realise opposition for all three features, unlike Icelandic. Although no distinction is made in the plural, the three singular persons are all distinct from each other and from the infinitive. Lack of syncretism between these forms means realisation of both a plus and a minus value for the

*speaker* as well as the *addressee* feature. This, together with the presence of a plural morpheme, leads to the realisation of full feature opposition.

Indeed, there is strong evidence that Old English allowed pro-drop (van Gelderen 2000:127/131, from *Beowulf*):

14. a. *namon þa to rede*  
       took then to council  
       'They took then to council'
- b. *þæt syðþan na ymb brotne ford brimliðende*  
       that since-then never on broad water-way seafarers  
       lade ne letton  
       passage not let  
       'that they after that never kept people from passing that water'.

In some of the Old English texts analysed by Berndt (1956), the rate of subject omission was as high as 43%. Also, note that example (13b) involves dropping the subject of an embedded clause, indicating that it is not a case of topic-drop (cf. footnote 2). Many such examples were also reported in Pogatscher (1901), cited in van Gelderen 2000.

Given our definition of richness, it is not surprising that Old English allowed pro-drop despite the fact that it failed to distinguish between *speaker* and *addressee* in the plural. This is because full feature opposition could still be realised by the distinction between each singular affix. However, this was no longer true in the Late Middle English period. By then, the infinitive and the plural affixes had converged, with consequent loss of opposition for the *singular* feature:

15. Middle English paradigm: Inf.: sing-**en** ('to sing')
- |     |                  |             |     |                 |
|-----|------------------|-------------|-----|-----------------|
| 1sg | sing- <b>e</b>   | +sp aad +sg | 1pl | sing- <b>en</b> |
| 2sg | sing- <b>est</b> | -sp +ad +sg | 2pl | sing- <b>en</b> |
| 3sg | sing- <b>eð</b>  | -sp -ad +sg | 3pl | sing- <b>en</b> |

This leads us to expect that pro-drop should no longer have been allowed in Middle English. Though empirical knowledge in this area is not uncontroversial, the situation seems promising.

Among the Middle English texts recently analysed by van Gelderen (2000) there are two works that are of particular interest to our discussion, namely the

*History of the Holy Rood Tree* (HRT) and a collection of texts known as the 'Katherine Group' (KG). These texts differ with regard to pro-drop in a manner which is crucial to our proposal: while pro-drop occurs "relatively freely" (van Gelderen 2000:145) in the KG text, there is no evidence of pro-drop in the HRT text. An example from the KG text is given below:

16. Cleopest þeo þing godes...  
 Call.2sg those things good  
 "Do you call those things good?"

On further inspection, we find that – while the HRT variety conforms to the paradigm in (15), the KG texts are a different matter as "[t]he Old English verbal system is remarkably well preserved in these texts" (Logan 1973:185). In particular, this variety still used "different vowel combinations with –þ for the plural" (van Gelderen 2000:174) indicating that the –sg value was still being spelt out and therefore opposition for the *singular* feature was realised. It is thus unsurprising that this variety allowed pro-drop.

#### 2.1.5 The Irish Paradigm

Modern Irish has been claimed to be problematic for morphological analyses that relate pro-drop to rich agreement systems at least since the work of McCloskey and Hale (1984). On the one hand, the Irish system does not seem to be particularly rich. As shown in (17), this language does not realise opposition for the *addressee* feature as it employs the same morpheme, namely –*eadh*, for 2nd person singular, 3rd person singular and 3rd person plural (adapted from Müller 2005:12):

17. Irish paradigm: Cond.: cuir-**f** ('put.COND')<sup>8</sup>
- |     |                     |             |     |                     |              |
|-----|---------------------|-------------|-----|---------------------|--------------|
| 1sg | chuirf- <b>inn</b>  | +sp aad +sg | 1pl | chuirf- <b>imis</b> | +sp aad -sg  |
| 2sg | chuirf- <b>ea</b>   | -sp +ad +sg | 2pl | chuirf- <b>eadh</b> | -sp aad βasg |
| 3sg | chuirf- <b>eadh</b> | -sp aad βsg | 3pl | chuirf- <b>eadh</b> | -sp aad βsg  |

Assuming the definition of richness I am proposing, the prediction that follows is that Irish should not allow pro-drop. Nevertheless, claims to the contrary have often been made. Examples such as (18) have been cited in support of such claims (from McCloskey and Hale 1984):



18. Chuirfinn isteach ar an phot sin  
 put.COND.1SG in on that job  
 'I would apply for that job'

However, the conclusion that (18) is an instance of pro-drop seems too simplistic. Unlike the classic pro-drop languages, Irish has the property that agreement morphology must be absent when the subject is overtly realised. Compare the examples in (19) with (18) above:

19. a. \*Chuirfinn mé isteach ar an phost sin  
 put.COND.1SG I in on that job  
 b. \*Chuirfeadh isteach ar an phost sin  
 put.COND in on that job  
 c. Chuirfeadh Eoghan isteach ar an phost sin  
 put.COND Owen in on that job

Note that example (19a) is ungrammatical even in contrastive contexts. In order to build a contrastive reading in Irish, the contrastive particle *-se* is attached to the verbal ending instead (Doron 1988).

The Irish elements have at least one other characteristic which is problematic for a pro-drop analysis. As shown in (20), they can occur in co-ordinated structures (from McCloskey and Hale 1984):

20. dá mbe-inn -se agus tu -sa ann  
 if be.COND.1SG -contr. and you -contr. there  
 'if you and I were there'

This property is typically associated with syntactic rather than morphological elements (see for example Zwicky 1977, 1985), a fact that should make us immediately suspicious of the status of the Irish items. Unsurprisingly, structures like (20) are impossible in pro-drop languages, whether the verbal agreement refers to the silent pronoun or to the whole co-ordination:

21. a. \*Se io e parli/parliamo Italian  
 if I and speak.2SG/1PL  
 b. \*An ego ke milisis/milisume Greek  
 if I and speak.2SG/1PL  
 'if you and I speak'

What these data seem to suggest is that the Irish elements are not instances of morphological agreement at all. Indeed, those proposals that have tried to analyse the Irish data in terms of agreement (e.g. McCloskey and Hale 1984; Legate 1999) inevitably had to appeal to additional stipulations in order to capture the complementary distribution between pronouns and the alleged agreement morphemes.

For this reason, a number of researchers have argued that the phenomenon found in Irish (and other Celtic languages) results from interface constraints that apply to the verb-subject sequence and have nothing to do with agreement or the licensing of empty pronouns (see Pranka 1983; Doron 1988; Ackema and Neeleman 2003 for cliticisation/allomorphy analyses). A major advantage of these proposals is that they provide a straightforward explanation for the complementarity shown in (18) and (19), since they take the two elements to be in an allomorphic relation which surfaces in specific circumstances. In particular, the analysis proposed by Ackema and Neeleman (2003) also captures the fact that this type of complementarity seems to be found only in OVS languages. The superior explanatory power of these proposals when compared to agreement analyses indicates that the similarities which Irish seems to share with pro-drop languages are only superficial. In this respect, the intuition behind the label 'pro-drop' seems to be rather accurate; pro-drop languages are languages that drop pronouns, not affixes. In conclusion, Irish is most probably not a pro-drop language and therefore it appears that the prediction stemming from the formal properties given in (17) is in fact borne out<sup>9</sup>.

Though a wider range of languages may have to be considered in order to test this hypothesis further, we can conclude that the idea of feature opposition can provide us with a satisfactory definition of "richness". In the next section I will look at how this can shed some light on the relationship between pro-drop and V-to-I movement.

#### *2.1.6 Explaining Koenenman and Neeleman's Generalisation*

The binary system introduced in section 2.1.2 has also been implemented by Koenenman (2000) in his investigation of V-to-I movement. Koenenman argues that V-to-I movement is triggered by those agreement paradigms which are sufficiently rich to realise all three features: *addressee*, *speaker*, *singular*. One such example is found in the Icelandic paradigm (cf. 8 above) repeated here for convenience as (22):

22. Icelandic paradigm; Inf: seg-**ja** ('to say')

1sg	seg- <b>i</b>	+sp aad +sg	1pl	seg- <b>jum</b>	+sp aad -sg
2sg	seg- <b>ir</b>	-sp aad +sg	2pl	seg- <b>ið</b>	-sp +ad -sg
3sg	seg- <b>ir</b>	-sp aad +sg	3pl	seg- <b>ja</b>	

Although the Icelandic paradigm does not realise full feature opposition, it is sufficiently rich to realise at least one value for each of the three features. In fact, the feature *singular* is present in both a positive and a negative value, as shown by the distinction between, for example, 1st person singular and 1st person plural. In the same vein, the two separate values for the *speaker* feature are also realised, given that - in the singular - a distinction is made between 1st person on the one hand and 2nd and 3rd person on the other. Finally, although the feature *addressee* only appears in one of the specified values, namely plus, it is nonetheless present due to the fact that the plural suffix - ið is compatible only with second person. It is the presence of all these three features that Koenenman (2000) claims to be responsible for V-to-I movement in Icelandic.

This contrasts with languages that do not realise all the three features. One such language is Halligdalen Norwegian. In this language, the only distinction we find is between the singular and the infinitival/default form (from Koenenman 2000:61):

23. Halligdalen Norwegian paradigm; Inf: kast-**æ** ('to sling')

1sg	kast- <b>a</b>	+sg	1pl	kast- <b>æ</b>
2sg	kast- <b>a</b>	+sg	2pl	kast- <b>æ</b>
3sg	kast- <b>a</b>	+sg	3pl	kast- <b>æ</b>

The -a affix is compatible only with singular contexts and is therefore specified as [+sg]. As to the plural, Halligdalen Norwegian resembles Middle English (cf. 15 above) in that it employs the default infinitival form. According to Koenenman, this prevents Halligdalen Norwegian from allowing V-to-I movement, a prediction which he shows to be borne out (note that Halligdalen Norwegian also disallows pro-drop, in accordance with our definition of richness).

If Koenenman's account of V-to-I movement is correct, an interesting corollary emerges: what is sufficiently rich for pro-drop will also be sufficiently rich for V-to-I. This follows if we consider that in order for a paradigm to realise full opposition it must necessarily realise all three features. Indeed, evidence of a near-

correlation between V-to-I and pro-drop has often been given in the literature<sup>10</sup> (see for example Kayne 1991, Platzack and Holmberg 1989, Roberts 1993, among others).

In particular, Koenenman and Neeleman (2001) categorise languages in relation to a "scale of inflection" which suggests that pro-drop languages are a subset of V-to-I languages:

#### 24. Scale of Inflection

Poor	Middle Class	Rich
No V to I / No pro-drop	V to I / No pro-drop	V to I / pro-drop

If the idea of feature opposition discussed in section 2.1.3 is correct, the generalisation in (24) can be explained: because the specification that allows pro-drop is a subset of the specification that triggers V-to-I movement, pro-drop languages will be a subset of the languages that display V-to-I. This seems to be an interesting side-effect of our definition of richness.

#### 2.1.7 Speculations on Opposition

Naturally, the view I have presented here raises the question of why feature opposition should be an important property of natural language. In particular, if lack of ambiguity is a desirable property, why do pro-drop languages not require realisation of a distinct morphological marker corresponding to each person (e.g. Italian cf. 6)? Although this would make the wrong prediction for Portuguese (cf. 9), it may seem the only viable strategy that would provide pro-drop languages with truly unambiguous paradigms. However, if we take seriously the idea that Person and Number features should be of a binary nature (see section 2.1.2), we must rule out any requirement resorting to the notion of 1st, 2nd, and 3rd person, since these are not primitive concepts. Of course, one could formulate a requirement in terms of binary features in such a way that it would apply to all three persons. However, such requirement would be entirely arbitrary, as the forms it would target do not constitute a natural class.

A more viable alternative would be to equate unambiguous paradigms to the instantiation of all possible feature combinations. Although this possibility would be compatible with feature binarity, it must nonetheless be rejected, as it seems empirically inadequate in more than one respect. Firstly, as we have seen in (6), a language like Italian is sufficiently rich to spell out six different person/number combinations but does not actually realise all possible combinations as it lacks the

inclusive/exclusive distinction in the first person (unlike – for example – Arabic, cf. footnote 4).

Moreover, this view necessarily implies that realising all possible feature combinations is a fundamental property of natural language, independently of pro-drop. What makes pro-drop languages different would be that they express this property on the verb rather than pronominally. This seems to be untrue as there are a number of languages which lack combinations of all features whether on the verb or otherwise. In Standard British English, for example, the combination [*-sp*, *+ad*, *+sg*] is altogether absent since the pronoun *you* does not spell out a singular/plural distinction. Though potentially ambiguous, utterances like ‘you saw him’ are not ill-formed. Note that, when looking at the pronominal system of English, we find that full feature opposition is indeed realised, despite the 2nd person ambiguity:

25. English pronominal paradigm

1sg	I	+sp aad +sg	1pl	we	+sp aad -sg
2sg	you	-sp +ad a sg	2pl	you	-sp +ad a sg
3sg	he/she	-sp -ad +sg	3pl	they	-sp -ad -sg

This might imply that our proposal could be extended to a more general – perhaps universal – requirement which may be formulated as follows: full feature opposition must be lexically realised. This would be independent of whether a language does so via pronouns or verbal agreement. If this is correct, pro-drop is simply a label we use when the latter case applies.

A question that might arise in response to this is why lack of ambiguity holds at the paradigm level rather than at the level of the utterance. In other words, why is it that a system would require morphological paradigms to be formally unambiguous when it is clear that no such requirement holds for some individual members of the paradigm (e.g. English *you*, Portuguese *-am*). Unfortunately, I do not have an answer to this but I suspect it may be related to some “minimal content” requirement that holds at the syntax/pragmatics interface. Very simplistically, ambiguity may be resolved by pragmatics provided that some formal content is provided upon which the speaker/hearer’s expectations can be grafted. This content need not be encoded in any given utterance, however. For example, it may be inherent in the mutual knowledge that speaker and hearer share. This may stem from the need to optimise computational effort which, in the absence of a minimal formal content, would perhaps be too great (indeed, it is difficult to imagine

how a language with neither pronouns nor verbal agreement would be able to cope). In essence, what I am proposing is that this minimal formal content takes the shape of full feature opposition<sup>11</sup>. As feature opposition is a property of morphological paradigms, it provides minimal content in the form of mutual formal knowledge between speaker and hearer.

As far as I can see, there is one more process through which the concept of unambiguous paradigm could be encoded. I will only consider this briefly as it suffers from the same empirical inadequacy as the one just mentioned, despite its appealing to a less strict requirement. This requirement would be the need to specify all occurrences of all Person/Number properties (i.e. *speaker*, *addressee*, *singular*). This is a looser requirement as it only expects specification iff a feature is present. Nevertheless, it does not provide a solution as it still disallows pro-drop languages to lack the same inclusive/exclusive distinction discussed above and it would predict that the ambiguity associated with English *you* should not exist.

Therefore, it seems that – among the linguistically definable candidates that could encode the idea of unambiguous paradigm – feature opposition is the only one that does not immediately fail on empirical adequacy.

#### 2.1.8 Interim

In sections 2.1.1-2.1.7 I have proposed a general definition of morphological richness with regard to verbal paradigms. I suggested that richness is encoded at the level of morphological specification via spell-out of what I have been calling feature opposition. Languages whose morphology distinguishes between a positive and a negative setting for the features *speaker*, *addressee* and *singular* qualify as morphologically rich while other languages do not. Although a wider range of languages needs to be examined before this definition might be considered as a possible universal, it provides a straightforward solution to traditionally recalcitrant cases such as German and Icelandic, as well as the different pro-drop status of Brazilian as opposed to European Portuguese. I have also shown that this definition sheds some light on the near correlation between pro-drop and V-to-I movement.

A consequence of this account is that pro-drop is not due to direct lexical differences between languages (pronouns may well share the same features across languages). Instead, it emerges from the combination of a (i) universal economy-type principle and (ii) language-specific properties of verbal – rather than pronominal – morphology. In the next sections I will discuss the implications that this has with regard to the Cascade-updating model.

## 2.2 Consequence for Cascade-updating

### 2.2.1 Transfer and Pro-drop

The principle in (3) has two fundamental properties: it is not subject to cross-linguistic variation and it reduces pro-drop to feature visibility. This has an important consequence for the Cascade-updating model: because Cascade-updating applies within the lexicon, anything encoded outside the lexical domain cannot possibly be affected. The only way in which a child acquiring, say, English could drop subject pronouns is if s/he ends up with a verbal paradigm that realises full feature opposition<sup>12</sup>. Clearly, this cannot obtain via experience since there is no evidence for such realisation in the target language. However, the question arises as to whether Cascade-updating might spread specified Person and Number properties over to a poor agreement language such as English, subsequently triggering the principle in (3). In the same vein, we must address the question of whether the opposite could obtain, namely the transfer of underspecified properties over to a rich agreement language, thus inhibiting the realisation of pro-drop.

For Cascade-updating to spread the (under)specifications of one language over to the other it is necessary that Paradigm Formation perceives the two sets of items as sharing the same paradigmatic properties. Below I consider how this would work in a hypothetical case of English-Italian bilingualism.

### 2.2.2 Morphological Transfer

As discussed in chapter 1 (section 2.2), the English verbal paradigm has a general form plus one more specified form, namely -s. Note that this latter form cannot have any specification for *speaker* or *addressee* since there is no evidence in the input which could trigger postulation of a contrast between these features. Indeed, it is far from clear what feature the -s morpheme does realise (see Koenenman 2000 for discussion). For the purposes of this discussion I will assume that it bears specification for the singular feature<sup>13</sup>:

26. English paradigm:inf.: walk

1sg	walk-Ø	1pl	walk-Ø
2sg	walk-Ø	2pl	walk-Ø
3sg	walks +sg	3pl	walk-Ø

Comparing the above with the Italian paradigm in (6) it is apparent that a language as morphologically poor as English cannot possibly affect a richly inflected language since this would involve transferring the lack of a property, an impossible outcome

within the Cascade-updating model. Note that, even if we were to assume that the English form *walk* bears some kind of feature – perhaps related to the infinitival nature of the item – transfer would still be ruled out. This is because the presence of such a feature, call it [-realised] (see section 2.1.3 above) would cause the English item to be perceived as carrying distinct paradigmatic properties compared to the Italian counterparts. This is illustrated below by (27) and (28) respectively, where the feature *dependent* indicates that the morpheme at issue is necessarily bound:

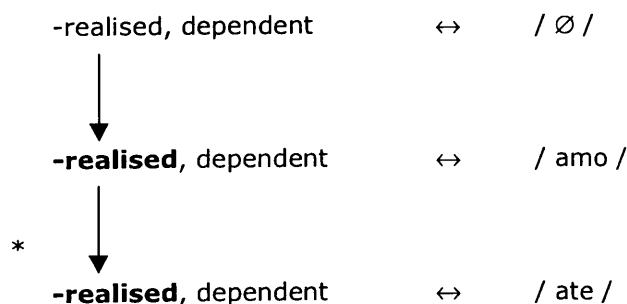
27. -realised, dependent

28. sp, ad, sg, dependent

At this point, a question arises as to whether the Italian elements could be affected at an earlier stage since, prior to the acquisition of the Person and Number features, they would only carry the *dependent* feature, thus potentially giving rise to transfer of the *-realised* feature.

Nevertheless, although it may be the case that this hypothetical earlier stage would trigger a transfer effect, this would not be sufficient to affect the pro-drop status of Italian. This is because transfer of the *-realised* feature to more than one Italian element would trigger a violation of the biuniqueness principle (cf. chapter 1, section 2.2). As discussed above, the poverty of the English paradigm follows from the fact that it only includes two forms. One of these, namely the zero-affix, supposedly bears the features in (27). Although transfer of these features to one of the Italian affixes is a logical possibility, transfer to more than one affix would necessarily result in one feature bundle being mapped onto two separate phonologies, consequently violating biuniqueness:

29.



In fact, the hypothetical stage whereby more than one Italian element would be carrying only a single *dependent* feature – a necessary condition for (29) to apply –



will also be a violation of biuniqueness. Therefore, a scenario where two Italian elements are postulated as mapping onto a single *dependent* feature must be ruled out for the same reason as (29).

The same violation of biuniqueness would result if transfer were to apply in the opposite direction, namely from Italian to English. Because English only has one unspecified morpheme, namely the zero-morpheme, even if the necessary conditions for transfer were to be met, this would only result in the transfer of one particular feature bundle, perhaps yielding a mapping whereby the English zero-morpheme would gain some specification. Nevertheless, this would not be sufficient to trigger pro-drop since the English paradigm would still be far from full-specification. Even so, it is far from obvious that such transfer should occur, as transfer of feature specification would be incompatible with the infinitival status that the English items must necessarily bear (see section 2.1.3 above). Though English children develop the third person marker rather late they do not have problems with the general cases, suggesting that – whatever feature they associate with infinitivals – is acquired rather early.

One final case that must be considered as a potential trigger for transfer regards the English third person. Unlike the infinitival, the English third person singular shares a feature specification with part of the Italian paradigm, namely [+sg]. This could be a potential trigger for Cascade-updating if there is any newly acquired feature that could be perceived as targeting [+sg]. If this were to occur, the English 3rd person morpheme could come to inherit richness from the Italian counterpart as a consequence of transfer.

Given Paradigm Formation, however, this does not seem to be a possibility. Looking at the items specified as [+sg] within the Italian paradigm, it is evident that there exists no feature that they all share besides [+sg] itself:

30.	1sg	parl- <b>o</b>	+sp aad +sg
	2sg	parl- <b>i</b>	-sp +ad +sg
	3sg	parl- <b>a</b>	-sp -ad +sg

For Cascade-updating to apply it would be necessary that +sg be associated with the presence of some other feature. However, given the situation in (30) no such feature exists and hence no triggering of Cascade-updating may occur.

### 2.2.3 Interim

I began our discussion by pointing out that the architecture of the system within which the Cascade-updating model operates is incompatible with transfer of a typological property such as pro-drop. I then addressed the question of whether feature specifications could be transferred instead, thus triggering pro-drop in an otherwise non-pro-drop language. In section 2.2.2 above I considered a hypothetical case of English-Italian bilingualism and argued that, although some specifications may conceivably be affected, transfer of paradigm richness is also incompatible with the Cascade-updating model.

In the sections that follow I will discuss three separate studies which have investigated the possibility of transferring pro-drop and I will suggest that they provide evidence in favour of the Cascade-updating model. Note that, in order to control for 'normal' pro-drop in child English (Radford 1990) we will only consider developmental patterns after the age of 2;5 (see also note 12).

## 2.3 Evidence from Bilingual Studies

### 2.3.1 Catalan-English

Juan-Garau and Perz-Vidal (2000) carried out a longitudinal study of a Catalan-English bilingual child called Andreu. Andreu was studied between the age of 1;3 and 4;2, (though no records were kept between the age of 3;0 and 3;1). The method of data collection was audio/video recordings and a parental diary.

The language pair being acquired by Andreu provides a good testing ground for the Cascade-updating model. Unlike English, Catalan is a pro-drop language with rich verbal agreement which realises full feature opposition:

#### 31. Catalan paradigm Inf.: cant-**ar** ('to sing')

1sg	cant- <b>o</b>	+sp aad +sg	1pl	cant- <b>em</b>	+sp aad -sg
2sg	cant- <b>es</b>	-sp +ad +sg	2pl	cant- <b>eu</b>	-sp +ad -sg
3sg	cant- <b>a</b>	-sp -ad +sg	3pl	cant- <b>e</b>	-sp -ad -sg

In a survey of Catalan utterances, Vallduvi (1987) reports that speakers drop pronominal subjects on average around 84% of the time, thus confirming that Catalan instantiates the principle in (3). According to the Cascade-updating model, however, this typological property cannot have any effect on bilingual development since it is not lexical in nature.

The subject of Juan-Garau and Perz-Vidal's study showed a strong preference for null-subjects in Catalan while consistently producing overt subjects in

English. As can be seen from the summary in table (32), the results of this study are strikingly similar to those presented in monolingual studies on English children (Ingham 1992, Valian 1990):

32. 1st Case Study: Comparison with Monolinguals<sup>14</sup>

	Age range	Language <sup>15</sup>	MLU	null subjects
<b>Bilingual</b>				
	3;2 – 3;3	Catalan		89.7%
		English	2.02	6.2%
	3;4	Catalan		75.0%
		English	~3.0 <sup>16</sup>	2.9%
<b>Monolingual</b>				
	2;5	--		
		English	2.61	9.0%
	2;5	--		
		English	3.57	7.0%

As expected, Andreu's English patterns closely with that of his monolingual peers as far as subject realisation is concerned. This is even more striking if we consider that Andreu's linguistic development was reported as being behind that of monolingual English children in a number of areas.

As a result of this, Juan-Garau and Perz-Vidal conclude that "the data available suggest the absence of any major influence of one language on the other" and that Andreu "appeared to follow separate routes in his syntactic development in each language". However, these statements follow only if we assume that pro-drop must necessarily provide an area of contact between the two languages. As argued in sections 1 and 1.1 above, this is not necessarily the case. The fact that Andreu used obligatory subjects in English at the same rate as English monolinguals is quite compatible with the idea that the two languages may influence each other. What the data do show, however, is that pro-drop is not an area of possible influence since the typological difference at issue is not lexically encoded.

### 2.3.2 Italian-English

Another case-study that is relevant to our discussion is that of Serratrice et al. (2004). This particular study investigated subject omission in the speech of Carlo, a bilingual child who was simultaneously acquiring Italian and English. The child was studied between the age of 1;10 and 4;6.

Just like Catalan, Italian differs from English in that it is a pro-drop language which instantiates the principle in (3). As shown in (6) above, Italian realises full feature opposition. In cases of English-Italian bilingualism, therefore, the Cascade-updating model makes the same prediction as for the English-Catalan case, namely that the rate of pro-drop in the English of English-Italian bilinguals should not differ from that of English monolinguals.

Serratrice et al. found that Carlo's Italian involves a strong preference for null-subjects while his English systematically features overt subjects. Once again, these findings are strikingly similar to those presented in monolingual studies on English children (Ingham 1992, Valian 1990). The results are summarised in the table in (33)<sup>17</sup>:

33. 2nd Case Study: Comparison with Monolinguals

	Age range	Language	MLU	null subjects
<b>Bilingual</b>	N/A <sup>18</sup>	Italian		51%
		English	2;0-3;0	7%
	N/A	Italian		59%
		English	3;0-4;0	3%
<b>Monolingual</b>				
	2;5	Italian		66%
		English	2;0-3;0	9%
	2;5	Italian		70%
		English	3;0-4;0	7%

For consistency, the English monolingual data were taken from Valian (1990), though Serratrice et al. (2004) compared Carlo's development with that of four English monolinguals from the CHILDES database (MacWhinney 2000). There is no considerable difference between the two sets of results.

As expected, the subject omission rate in Carlo's English converges with that of his monolingual peers, indicating that there is no interference from Italian. However, looking at the group data for the Italian monolinguals, it appears that Carlo realises more subjects than his monolingual Italian peers, thus raising the question of whether transfer did occur in the opposite direction, namely from English to Italian. Indeed, Serratrice et al. conclude that "crosslinguistic influence" (2004:198) has taken place. In addressing this point, however, it must first be noted that this domain presents a considerable amount of individual variation across

children. This is apparent when looking at the Italian monolinguals individually (from Serratrice et al. 2004):

34. Rate of Subject-drop by child

MLU	Diana	Guglielmo	Martina	Raffaello	Rosa	Viola
2;0-3;0	75%	70%	59%	77%	59%	69%
3;0-4;0	N/A <sup>19</sup>	70%	N/A	74%	64%	N/A
>4;0	65%	66%	N/A	N/A	N/A	N/A

Although Carlo's use of overt subjects is at the lower end of the spectrum, it does not differ significantly from that of some monolingual Italian children. A chi-square test revealed no significant difference between Carlo and Rosa at MLU 2;0-3;0 ( $\chi^2=3.782$ ,  $df=1$ ,  $p=0.052$ ) as well as MLU 3;0-4;0 ( $\chi^2=2.721$ ,  $df=1$ ,  $p=0.099$ ). No significant difference was also found between Carlo and Guglielmo at MLU >4 ( $\chi^2=2.794$ ,  $df=1$ ,  $p=0.95$ ). However, the study carried out by Serratrice et al. does not contain much monolingual data at MLU >4;0, as only 2 monolingual children out of the 6 who took part in the study provided data for this stage. Moreover, one of these, namely Guglielmo, only provided a mere 96 utterances. For these reasons, I conducted further investigation of the CHILDES database by analysing files from the Antelmi corpus (Antelmi 1992).

This corpus consisted of 7 files containing the utterances of Camilla, a monolingual Italian child. Four of the seven files contained utterances at MLU >4, providing a total of 620 tensed utterances. Of these, 285 involved an overt subject either in the form of a pronoun, demonstrative or some other NP. The coding involved marking each utterance as either *overt*, *NP* or *drop*. Sentences repeated twice within the same utterance were not coded. Also, following the study by Serratrice et al., imperatives were not coded.

The ratio of subject drop in Camilla's corpus amounts to 54%, which is lower than that of Carlo's, though not significantly lower. Consequently, it appears that – although Carlo's subject omission rate is at the lower end of the spectrum – it does not constitute evidence for transfer.

In conclusion, the fact that Carlo happened to use a relatively high number of overt subjects in Italian does not present a potential argument for transfer from English to Italian since his performance does not differ significantly from that of some of the monolingual controls. Moreover, further investigation indicates that there is at least one monolingual Italian child (Antelmi 1992) whose rate of overt subjects in Italian is as low as that of Carlo's at MLU >4;0. Finally, no evidence of

transfer in the opposite direction (i.e. from Italian to English) was found. Carlo's case is therefore in line with the idea that pro-drop is immune to transfer, as expected within the Cascade-update model.

### 2.3.3 Inuktitut-English

The final work that will be discussed with regard to pro-drop is that of Zwanziger et al (2005). This study involved the videotaping of six Inuktitut-English bilinguals over a period of one year (each child was recorded for a total of 4/6 times). The age of the children ranged from between 1;8 and 2;11 at the beginning of the study to between 2;6 and 3;9 when the study was completed.

Inuktitut is very different from the languages discussed so far. Unlike Catalan and Italian, it is a member of the Eskimo-Aleut family. Moreover, it is a polysynthetic (or incorporating) language which has basic SOV order and realises overt object as well as subject agreement. Nevertheless, Inuktitut shares one important feature with the Romance languages discussed above, namely its pro-drop nature. Despite other differences, it seems that Inuktitut allows pro-drop for the same reason that Catalan and Italian do, namely the realisation of feature opposition (adapted from Fortescue 1984):

#### 35. Inuktitut paradigm; Inf.: urni-**ssa** ('to come')

1sg urni- <b>vunga</b>	+sp aad +sg	1pl urni- <b>vugut</b>	+sp aad -sg
2sg urni- <b>vutit</b>	-sp +ad +sg	2pl urni- <b>vusi</b>	-sp +ad -sg
3sg urni- <b>vug</b>	-sp -ad +sg	3pl urni- <b>pput</b>	-sp -ad -sg

If this is the case, the prediction that follows in conjunction with the Cascade-updating model is that the pro-drop in English-Inuktitut bilinguals should mirror that of monolinguals.

Indeed, Zwanziger et al. (2005) reported that all the six children that took part in their study consistently dropped subjects in Inuktitut but realised them in English. As can be seen from the table below, the subject omission rate in English patterns closely with that of monolingual English children:

## 36. Group Study: Comparison with English Monolinguals

	Age range	Language	null subjects
<b>Bilingual</b>			
	2;1-2;6	Inuktitut	84%
		English	23%
	2;7-3;0	Inuktitut	82%
		English	3%
	3;1-3;6	Inuktitut	85%
		English	<1%
	3;7-3;9	Inuktitut	89%
		English	7%
<b>Monolingual</b>			
	1;10-2;2	English	28%
	2;3-2;8	English	9%
	2;5-2;7	English	7%
	2;6-2-8	English	3%

As expected, these results suggest that no transfer has taken place from Inuktitut to English. However, in order to test the possibility of transfer in the opposite direction (i.e. English to Inuktitut), a comparison with data from monolingual Inuktitut children is called for. Unfortunately, studies on Inuktitut monolingual children do not provide data that can be straightforwardly compared with the results in (19). The only study on the acquisition of Inuktitut which discusses subject-drop is that of Allen and Schröder (2003). However, this study does not divide the results into groups either by MLU or by age and only gives subject-drop rates as totals<sup>20</sup>. Nevertheless, for the purposes of our discussion it may be sufficient to note that monolingual Inuktitut children between the ages of 2;0 and 3;6 were reported to have an average rate of subject-drop of 84%, which closely reflects that of their bilingual peers (cf. 36 above). This figure, together with the results reported in (36), provides strong evidence in favour of the hypothesis that the 'property' of pro-drop cannot be transferred, as predicted by the Cascade-update model.

On the other hand, Zwanziger et al. conclude that "previously identified triggers for crosslinguistic influence do not operate universally" (2005:908). As argued above, however, the idea that pro-drop is a potential trigger for crosslinguistic influence is based on the flawed assumption that transfer can be triggered by any typological difference. Assuming the Cascade-updating model allows us to limit the potential triggers for transfer to typological differences that

are lexically encoded (though these are also subject to restrictions, see chapter 3). The results in (36) are fully compatible with this view.

#### 2.3.4 Interim

In this section it was argued that – within the Cascade-updating model – a typological difference such as pro-drop does not present a potential case for transfer since it is not a lexically encoded property. With this in mind, we looked at three separate studies which investigated children acquiring language pairs that involved a pro-drop and a non pro-drop language. All of these studies provided evidence that no transfer of pro-drop had taken place.

In the sections that follow I will discuss another phenomenon whose triggers appear to lie outside the domain of Cascade-updating, namely Root-Infinitives.

### 3. Root Infinitives

#### 3.1 The Data

It has long been noted that children go through a stage when they produce main clause verbs in the infinitival form. This phenomenon is known as Root Infinitives (henceforth RIs) and has been reported for a number of languages:

- |     |  |                            |
|-----|--|----------------------------|
| 37. | a. Der ikke vaere<br>it not be                                 | Danish (Wexler, 1994)      |
|     | b. Papa shoenen wassen<br>Daddy shoes wash.INF                 | Dutch (Wexler, 1994)       |
|     | c. She drink apple juice                                       | English (Vainikka, 1994)   |
|     | d. Michel dormer<br>Michel sleep.INF                           | French (Pierce, 1992)      |
|     | e. Thorstn das haben<br>Thostn this have.INF                   | German (Wexler, 1994)      |
|     | f. Jag också hoppa där a där<br>I also hop.INF there and there | Swedish (Santelmann, 1995) |

In earlier work (e.g. Radford 1990) this phenomenon was taken to result from the child's grammar lacking functional categories such as Infl, thus forcing the verb to surface in its bare form. Although appealing as far as the English data are concerned, this hypothesis has been considered problematic at least since the work of Wexler's (1994) which showed its inadequacy at tackling languages other than



English. Firstly, as is evident from the examples in (37), the RI phenomenon does not simply represent the use of 'stripped' verbal roots since it occurs in conjunction with infinitival marking (e.g. 37b,d,e,f). Most importantly, it has been argued that there does not exist an RI stage as such, since RIs co-occur with finite forms (see for example Pierce 1992). If this position is correct<sup>21</sup>, the "maturation of Inflection" hypothesis is considerably weakened. A potential objection to this conclusion, however, is that the finite verbs that children produce at this stage may in fact be unanalysed chunks. On closer inspection, this view seems unlikely to be correct since – at the same stage – children consistently use the correct inflection when marking finiteness (Wexler 1994).

Another striking property of RIs is that they do not seem to show up in positions where infinitives are disallowed in the target language. For example, Weverink (1990, reported in Wexler 1994) observed that while Dutch children place finite verbs in second position, they repeatedly produce RIs clause finally. This is consistent with the SOV/V2 nature of the target language (see also Pierce 1992 and Wexler 1994 for similar data from French and German respectively). Compare (38) and (39) (adapted from Wexler 1994:316-317):

38. *Finite forms*

- a. ik pak        `t op  
   I pack.1SG it up
- b. groot is e bal  
   big    is    ball
- c. wou        Tobiasje hebben  
   wanted T.        have.INF

39. *Non-finite forms*

- a. pappa nieuwe scooter kopen  
   daddy new    scooter buy.INF
- b. ik ook lezen  
   I also read-INF
- c. pappa kranten        weg    doen  
   daddy newspapers away do.INF

In view of these data, it seems that finite verbs in child language should not be treated as unanalysed chunks. Notice that this changes the focus of the problem;

the question to be asked is no longer "why do children produce RIs" but rather "why are finite verbs optional in child language".

In addressing this question, it is necessary to consider the fact that RIs are not a crosslinguistic phenomenon. It has often been observed that RIs are rather rare in some languages (Guasti 1994, Cipriani et al. 1989, Sano 1995, among others). The table<sup>22</sup> in (40) presents some data in order to illustrate this typological variation (adapted from Hoekstra and Hyams 1998)<sup>23</sup>:

40. Root Infinitives in child language<sup>24</sup>

Lang	Child	age	RIs	Lang	Child	age	%RI
Italian	Diana	2;0	0	French	Nath.	1;7- 2;11	76%
	Martina	1;11 2;1	16% 4%		Daniel	1;5-2;5	60%
	Paola	2;0-2;5	7%	English	Eve	1;6- 1;10	78%
	Daniele	1;7-2;6	8%		Adam	2;3 3;0	- 81%
	Mass.	1;7-2;6	6%		Nina	2;4 2;5	- 75%
	Gabriele	1;7-2;6	7%		Freja	1;11- 2;0	38%
	Orietta	1;7-2;6	5%	Swedish	Tor	1;11- 2;2	56%
	Elisab.	1;7-2;5	10%		Embla	1;8- 1;10	61%
	Franc.	1;9-2;5	5%	German	S	2;1 2;2	53% 40%
	Damariz	2;6-2;8	5%		Laura	1;8-2;1	36%
Spanish	Juan	1;7-2;0 2;1-2;4	12% 10%	Dutch	Tobias	1;10- 1;11	36%
					Fedra	1;10- 2;1	26%
Catalan	Guillem	1;11- 2;6	3%				
	Marti	2;0-2;5	3%				

An interesting correlation emerges from the data in (40); the languages that show strong RI effects are non-pro-drop languages. As far as pro-drop languages are concerned, RIs seem to be extremely rare.

A number of hypotheses have been proposed in an attempt to capture this correlation. In particular, two traditions have emerged from the generative literature; the "omission" views pioneered by Wexler (1990, 1994, 1998, also Schutze and Wexler 1996) and the "interface" views of Hyams (1996) and Hoekstra and Hyams (1995, 1998). I will briefly outline them below<sup>25</sup>.

### **3.2 Infinitives, Agreement and Pro-drop**

According to Wexler (1998), rich agreement systems only need checking of one D feature, which he takes to be situated in T under the split-I hypothesis of Pollock (1989). Poor agreement systems, on the other hand, need checking of two D features, an assumption derived from the fact that these languages require an overt DP specifier.

The non-pro-drop/RI correlation is then derived by further postulating that children's grammar is constrained to the checking of only one feature within a single derivation (i.e. the Unique Checking Constraint). Children acquiring a language which requires overt DPs are therefore faced with a dilemma: leaving a D feature unchecked, and thus producing an ungrammatical derivation, or rescuing the derivation by inserting an infinitive. When the latter strategy is applied, RIs will obtain. Because children acquiring rich agreement systems are never faced with such a situation, RIs are correctly predicted to be lacking in these languages.

A different position is taken by Hyams (1996) and Hoekstra and Hyams (1995, 1998) who developed a number of related proposals where they suggest that the attested distribution of RIs follows if we assume that child language lacks Number specification. The lack of RIs in richly inflected languages is then expected since these languages can distinguish between finite/non-finite forms by independent means (i.e. Person and Tense features). Poorly inflected languages, on the other hand, often rely on a Number feature in order to establish this distinction. The RI effect in poorly inflected languages is then taken to arise from the child's inability to grammatically express finiteness. Because of this deficiency, children will produce non-finite verbs and rely on the discourse in order to license their interpretation (a view that echoes Grodzinsky and Reinhart's (1993) idea with regard to DPBE, see chapter 3 section 2.1).



the Cascade-updating model in conjunction with the richness account discussed in section 2.1 is incompatible with the occurrence of transfer effects within the domain of RIs. A number of studies have been presented in the literature with regard to the possibility of transferring RIs. I will outline each of them in turn.

### 3.3 Evidence from Bilingual Studies

#### 3.3.1 German-Italian

Berger-Morales and Salustri (2003) discuss a case-study of a German-Italian bilingual child called Leo who was recorded on average twice a month between the age of 2;0 and 2;7. Leo was reported as having been living in Italy with his German-speaking mother and Italian-speaking father who employed the principle of 'one parent one language'.

The language pair being acquired by Leo provides a good testing ground for the Cascade-updating model. Unlike Italian, German is a language that typically displays robust RI effects, as shown in (40) above. However, according to the principles in (41), this is not directly encoded in the lexicon but follows from independent properties, namely the presence/absence of feature opposition within the agreement system of each language. In the context of the Cascade-updating model this means that the two languages cannot have any effect on each other in cases of bilingual development since their typological difference is not lexical in nature.

Similarly to monolingual German children, Leo's speech was reported to contain RI structures, an example of which is given below (from Berger-Morales, Salustri and Gilkerson 2003:4):

42. Heidi gucken. (Leo 2;1)  
 Heidi watch.INF  
 "(I want to) watch Heidi."

In fact, the majority of the verbs Leo produced in this language were non-finite. This differs strikingly with data from Leo's Italian:

43. RI rate in German/Italian Case Study: Leo

Age Range	Language	RIs
2;0 – 2;4	German	81%
2;6 – 2;7	Italian	7%

The percentage of RIs in Leo's Italian mirrors that of some of the monolingual Italian children reported in the literature. At 7%, Leo's RI rate is identical to that of Paola and Gabriele and not far from that of other monolingual Italian children (cf. table in 40). As to the German data, RIs are clearly very high in Leo's speech. Nevertheless, this did not influence Leo's Italian and it is therefore in accordance with the Cascade-updating model.

A couple of comments are in order. The amount of RIs in Leo's German seems disproportionately high, especially when compared to the 53% rate reported in Guasti (1994). Since Italian has a very low rate of RIs, the unusually high number of RIs in Leo's German does not seem to be imputable to transfer from Italian. Moreover, I do not think that a theoretical explanation is called for. More plausibly, it may be the case that the apparent discrepancy between Leo's German and that of his monolingual peers is due to the small amount of data collected by Berger-Morales and Salustri (2003). In fact, only 63 instances of root clauses featured in Leo's recordings between age 2;0 and 2;4 as opposed to the 398 clauses upon which Guasti's calculations were based (collected during the comparable stage that runs between age 1;11 and 2;4). As the data in Berger-Morales and Salustri only amounted to a mere 20% of the data analysed by Guasti (1994), it may be the case that a comparison is difficult to draw. This suspicion is confirmed when we look, for example, at the monolingual case discussed in Berger-Morales and Salustri (2003). This included 77 root clauses, which is slightly higher than the amount collected for Leo's German and involved a lower percentage of RIs, at 67.5% (data from the Nijmegen corpus available from the CHILDES database, MacWhinney and Snow 1990, MacWhinney 2000).

In conclusion, this particular study may not be as informative as we would like with regard to the German data. However, the fact remains that the percentage of RIs in Leo's Italian was comparable to that of Italian monolinguals indicating that no transfer had taken place from German to Italian. In the next section I will look at a study that provides a higher amount of data and thus a potentially more fit comparison with the monolingual case in Guasti (1994).

### *3.3.2 German-Italian 2*

Hulk and Müller (2000) carried out a longitudinal study of another German-Italian bilingual child called Carlotta who was studied between the age of 1;8 and 3;1. The method of data collection was video recordings. As this study carried on for much longer than the one reported in Berger-Morales and Salustri (2003), a much larger

amount of data was obtained. For this reason, the results might be considered as providing a more reliable comparison with monolinguals.

The percentage of RIs in Carlotta's Italian was found to be very low. On the other hand, her German involved a high rate of RIs:

44. RI rate in German/Italian Case Study: Carlotta

Age	Language	RIs
2;3	German	42%
2;4	Italian	4.8%

The results in (44) pattern closely with the findings reported in monolingual studies (cf. 40 above). The RI rate in Carlotta's German is within the range reported by Guasti for S. (56% at age 2;1 and 40% at age 2;2). Similarly, the RI rate in Carlotta's Italian is virtually identical to that of Orietta and Francesco (both at 5%) as reported in Shaeffer (1990).

In sum, the results in (44) indicate that, as far as RIs are concerned, no transfer occurred during Carlotta's linguistic development. This is entirely compatible with the Cascade-updating model.

### 3.3.3 An Alternative Analysis

Hulk and Müller (2000) offer an analysis which could also explain the lack of transfer effects in Carlotta's case. According to their model, for transfer effects to occur the phenomenon at issue (RIs in this case) must:

45. a. involve the interface between syntax and pragmatics in the C domain.
- b. instantiate some overlap of the two systems at the surface level.

In setting the condition in (45a) Hulk and Müller follow Platzack (1999) in assuming that the C domain offers a particularly problematic area of acquisition. They relate this to the idea that the C domain interfaces between core grammar and other cognitive systems, as suggested by its role in expressing cross-modular information such as question formation, topicalisation and focus. Among other things, this is taken to supply a contact point between the two languages, potentially leading to transfer effects.

The condition in (45b), on the other hand, refers to potential ambiguity in the child's input. This condition claims that transfer effects may result when (i) language A has a surface construction which can be associated with more than one

structural analysis (one of which diverges from the target) and (ii) language B provides potentially reinforcing input for the non-target analysis. When this occurs, the two languages are said to overlap at the surface level. According to Hulk and Müller, this leads to crosslinguistic influence from language B to language A with consequent emergence of transfer effects in language A, provided that condition (45a) is also met.

With regard to the causes of RIs, Hulk and Müller (2000) assume Hoekstra and Hyams' (1998) model. As this model relates the acquisition of finiteness to the syntax/pragmatics interface, the condition in (45a) is automatically met. However, Hulk and Müller (2000) argue that the second condition cannot be fulfilled since no overlap exists between the two languages in this domain.

In relation to German/Italian case, this can be understood as follows. Firstly, consider that Italian does allow main clause infinitives when certain discourse conditions are met (from Hulk and Müller 2000:232, Rizzi 1994:375):

46.   a. Che cosa dire     in questi casi?  
          What     say.INF in these cases  
      b. partire     immediatamente!  
          Leave.INF immediately

This type of input may reinforce the child's default analysis that RIs may be discourse licensed in main clauses (cf. Hoekstra and Hyams 1998). According to Hulk and Müller, however, for transfer to take place it is necessary that the other L1 provide a surface overlap in this domain. Such overlap would be instantiated if, for example, adult German allowed infinitivals in main clauses more often than Italian does. Transfer would then be predicted to occur from German to Italian. In the same vein, if German allowed main clause infinitives less often than Italian the opposite prediction would follow. Note that this view relies on the concept of overlap at surface level. If German happened to allow main clause infinitives more/less often than Italian then transfer would be expected to occur, regardless of what the underlying reasons for this difference might be. The idea behind this analysis is that the child takes the linguistic input from one language and uses it as positive evidence in order to set a parameter in the other L1.

In the case of RIs, Hulk and Müller (2000) claim that no overlap obtains since German also allows main-clause infinitives within limited contexts (e.g. expressing a rhetorical function):



47. a. niet doen!  
not do.INF  
b. wie diese schliessen?!  
how these close

Because main-clause infinitives are marginally possible in both target languages, no surface overlap obtains and the results of their study are therefore expected (cf. 44). Note that, as Unsworth (2003) points out, in cases where the two languages do not differ it seems odd to talk about a lack of overlap since such cases seem to be instances of total overlap. Therefore, the notion of overlap as used by Hulk and Müller (2000) seems to necessarily imply partial overlap. Indeed, if we are to interpret (45b) as requiring a partial surface overlap between the two languages, the German/Italian case does not seem to qualify and therefore transfer incorrectly ruled out.

However, an important consequence of this model is that transfer of RIs may actually be predicted for some language pairs (see also Unsworth, 2003). On the other hand, the Cascade-updating model predicts that transfer of RIs should be ruled out crosslinguistically.

As far as the Cascade-updating model is concerned, RIs should never present a case for transfer since their presence results from the interaction of universal conditions and is therefore outside the domain within which Cascade-updating applies. However, a decisive factor on which Hulk and Müller's proposal depends in order to rule out transfer in the German/Italian case is that the two target languages do not differ with regard to the licensing of infinitival main clauses. They explicitly point this out: "what is crucial here is that the languages do not differ in this respect" (2000:232).

The situation seems different when we consider the case of a bilingual child simultaneously acquiring English and a non-RI language. The fact that English does not overtly distinguish between infinitives and finite root declaratives may give rise to the surface ambiguity necessary to fulfil the condition on surface overlap (i.e. (45b)). Although the child may eventually hear other cues suggesting that English main clauses do not actually license infinitives, the fact that Hulk and Müller's proposal relies on surface ambiguity means that – at least for a while – English main clause verbs must be potentially ambiguous for they can be analysed as either (i) finite verbs containing a phonologically empty affix or (ii) main-clause infinitives. This mirrors Hulk and Müller's own analysis of object-drop whereby they argue that Romance clitic structures are ambiguous between (i) chain structures formed by a

moved element plus a trace and (ii) object-drop structures where the object position is empty<sup>26</sup>. It is this type of ambiguity that they argue provides surface overlap.

Consequently, in bilingual cases when a child is acquiring English together with a language like Italian, which – as we saw in (46) – allows main-clause infinitives only in some specific discourse situations, the necessary overlap would be established. In such cases, the English input is expected to reinforce the child's default setting that allows RIs to be discourse licensed, thus leading to the prediction that transfer will occur from English to Italian, but not the other way around. As we shall see below, this prediction is not borne out.

### *3.3.4 English-Spanish*

Castro and Gavruseva (2003) carried out a longitudinal study that might help us test the different predictions which distinguish the Cascade-updating model from the interface model of Hulk and Müller (2000). This study involved searching the Deuchar corpus (Deuchar and Quay 2000) from the CHILDES database (MacWhinney 2000) in order to collect instances of RIs. The corpus contains data from a Spanish/English bilingual child, Mina, from age of 1;3 to age 2;6. The method of data collection was audio- and video-taping.

The language pair being acquired by Mina provides a good testing ground for the two competing hypotheses. As we saw from the summary in (40), Spanish and English differ greatly with regard to the amount of RIs reported in monolingual studies. In this respect, this particular language pair is no more informative than the one we have discussed above (i.e. German/Italian). However, English is unlike German in that it does not overtly mark infinitives and may therefore provide a surface ambiguity for the learner. This, according to Hulk and Müller's (2000) analysis, is of particular importance within the context of bilingualism. If a child acquiring English also happens to be simultaneously acquiring a language which allows main clause infinitives only in restricted contexts (cf. 46, 47), transfer is predicted to occur from the former language to the latter.

As shown in (48) below, Spanish seems to be such a language as it allows main clause infinitives in some specific contexts, just like German and Italian (from Etxepare and Grohmann 2005):

48. a. Yo ir a la fiesta ayer?!  
I go.INF to the party yesterday  
'Me go to that party?'

b. Yo fregar los platos otra vez?! Ni hablar!

I do.INF the dishes again no say

'Me do the dishes again?! No way!'

Consequently, the prediction that follows from Hulk and Müller's hypothesis is that Mina's Spanish should involve a number of RIs higher than that reported for monolingual Spanish children. According to the Cascade-updating model, on the other hand, different language pairs are expected to behave similarly since RIs are outside the lexical domain, a characteristic that grants them immunity to transfer effects. As Spanish and English differ with regard to richness of agreement, the different rate of RIs we find in these languages seems to follow from the principles in (41) in the same manner as for German and Italian. Consequently, no transfer of RIs is expected for either language pair.

Castro and Gavrusseva (2003) reported a very high rate of RIs in Mina's English but found that the phenomenon was virtually absent in her Spanish:

49. RI rate in English/Spanish Case Study: Mina

Age Range	Language	RIs
1;9 – 2;6	English	81%
1;9 – 2;6	Spanish	2%

The results in (49) are strikingly similar to those reported in monolingual studies (cf. 40 above). The RI rate in Mina's English is exactly the same as that of Adam's, while her Spanish data included even less RIs than the monolingual studies reported in the literature, though she is rather close to Damariz's performance (5%) as reported in Grinstead (1994).

It must be pointed out that the count in Castro and Gavrusseva's study was done by following Hoekstra and Hyams' standards, which makes their results particularly comparable with the monolingual findings reported in the table in (40) (see Castro and Gavrusseva 2003:181 for details).

In conclusion, it appears that the RI phenomenon is resistant to transfer effects irrespectively of the language pair at issue, contrary to the prediction made by Hulk and Müller's interface hypothesis. On the other hand, the explanation offered by the Cascade-updating model does not rely on language-specific properties but rather on a typological split, namely presence versus absence of feature opposition (i.e. rich agreement). On this view, RI rates in bilingual children

should pattern with those of monolinguals for all language pairs. The results in (49) appear to be in line with this claim.

### 3.3.5 What about AgrS?

Recall that the Cascade-updating model can offer an explanation to the data presented above provided that the different RI rates attested across languages are not a consequence of direct lexical variation. Although the idea that the distribution of RIs follows from some principles akin to (41) has been extensively argued for in the literature (Hyams 1996; Hoekstra and Hyams 1995, 1998; Wexler 1990, 1994, 1998; Schutze and Wexler 1996), Rizzi (1993/1994) takes a different approach by suggesting that the correlation between pro-drop and RIs is in fact indirect causation.

Building on Belletti (1990), he proposes that pro-drop serves as a cue to the child in order to associate infinitivals with a *strong* AgrS feature. Once this value has been set, infinitives will obligatorily raise to T. On the other hand, languages that disallow pro-drop will set the AgrS feature as *weak* by default and will therefore have no raising requirement. He further assumes that immature grammars do not specify that root clauses must be CPs. In principle, this allows children to produce VPs/Inf(itive)P(hrase)s (cf. Kayne 1991) as root clauses. The attested empirical generalisation can thus be derived: children acquiring non pro-drop languages will allow structures where the verb is base-generated in some position lower than T, thus producing RIs. In languages like Italian, however, infinitives must necessarily raise due to the strength of AgrS and therefore bare VPs containing infinitivals are independently ruled out.

The idea that Italian infinitives have different raising requirements from the same elements in languages like English and French finds support in the contrast in (50) (from Rizzi 1993/1994:385-386) together with the auxiliary assumption that NegP is universally located above TP (cf. Zanuttini 1991):

- |     |                                |                |
|-----|--------------------------------|----------------|
| 50. | a. *to read not the book       | <i>English</i> |
|     | a'. to not read the book       |                |
|     | b. *ne lire pas le livre       | <i>French</i>  |
|     | b'. ne pas lire le livre       |                |
|     | c. non leggere piu' il libro   | <i>Italian</i> |
|     | c'. *non piu' leggere il libro |                |

Within a representationally modular architecture, the strong feature putatively associated with infinitivals would be of a lexical nature. In the case of a bilingual child acquiring – for example – Italian and German, it is at least conceivable that the feature value *strong* might end up being associated with the German infinitivals<sup>27</sup> as a result of Cascade-updating (provided that the other requirements for the application of Cascade-updating are also met). This would lead to the undesirable prediction that the RI rate in bilinguals should differ from that of their monolingual peers. Given that the studies reported in the literature have found no transfer of RIs, we must therefore conclude that Rizzi's view is potentially incompatible with an explanation in terms of Cascade-updating.

Having said that, it must be pointed out that Rizzi's truncation account is not without problems. As mentioned above, in order to correctly capture the distributional differences in (50), Rizzi follows Zanuttini (1991) in assuming that NegP is universally located above TP. A consequence of this analysis is that RIs are predicted to be absent in negative structures, since NegP cannot be projected unless TP is also projected. Whilst this prediction is supported by data from child Dutch (Hoekstra and Jordens 1994), Wexler (1994) discusses data from Pierce (1992) which indicate that negated RIs are quite common in child French, thus presenting a potential problem for Rizzi's view.

Nevertheless, the truncation hypothesis derives a number of empirical generalisations with respect to RIs and it therefore has a lot of explanatory potential. These include the non-occurrence of wh-RIs as well as the lack of subject clitics in French RI constructions, both of which are predicted by the assumption that RI structures lack a CP layer where wh-elements and subject clitics would be located. More research is therefore needed in order to establish whether the truncation view can be reconciled with the Cascade-updating model. As things stand now, the Cascade-updating model can straightforwardly account for the lack of transfer effects in RIs provided that we also assume one of the two views sketched out in section 3.2 as our auxiliary hypothesis.

#### **4. Concluding Remarks**

At the beginning of this chapter I argued that those typological characteristics which are encoded outside the lexical domain must necessarily be unreachable by Cascade-updating and therefore immune to transfer. I then went on to present two specific phenomena which I argued to be rooted outside the lexical domain, namely pro-drop and Root-Infinitives. Experimental evidence concerning these phenomena

indicates that bilingual children behave similarly to their monolingual peers, thus suggesting absence of transfer as expected within the Cascade-updating model.

In relation to Root-infinitives, I went on to consider an alternative model which is also compatible with some of the experimental results. I then argued that this model must be dispreferred, as it predicts that transfer of RIs is dependent on the specific language pair, contrary to fact.

Besides providing further support for the Cascade-updating model, the findings discussed in this chapter are also a reminder that the lack of transfer effects within certain typological differences should not be taken merely as evidence that the two languages are developing separately. On the contrary, these situations may be giving us important information with regard to the underlying conditions that make transfer effects possible.

## NOTES

---

<sup>1</sup> The term “pro-drop” is meant to exclude so-called “radical pro-drop” languages (e.g. Chinese, Japanese). Although these also allow the omission of arguments, the distributional differences that we find between radical pro-drop and agreement-based pro-drop indicate that we are dealing with two separate phenomena (e.g. Tomioka 2003). Moreover, Neeleman and Szendroi (2005) have convincingly argued that radical pro-drop is conditioned by pronominal – rather than verbal – morphology, thus explaining why generalisations that have tried to collapse the two phenomena have been so unsuccessful (e.g. Jaeggli and Safir 1989; Speas 1993, 2006). I will disregard radical pro-drop throughout.

<sup>2</sup> It has often been shown that some Germanic languages allow omission of a pronoun when this is discourse given and has undergone movement to a sentence-initial position. E.g. Dutch (from Neeleman and Szendroi 2005:3):

- (i)  $O_1$  ken ik  $t_1$  niet  
    know I not  
    'I don't know *pro*'  
(ii)  $O_1$  ken  $t_1$  hem niet.  
      know him not  
      '*pro* don't know him'

However, this type of pronoun omission (usually referred to as "topic-drop") is a clearly distinct phenomenon as – unlike pro-drop – it is dependent on movement and affects both subjects as well as objects.

<sup>3</sup> This use of the term 'paradigm' might create some confusion. So far I used the term in a theory-specific sense in order to refer to the concept of paradigmatic relations within the lexicon, as defined in chapter 1. In this section, however, I will be using the same term in its more traditional sense when discussing morphological relations between affixes. Which of the two uses I am referring to will be made clear by the context.

<sup>4</sup> Note that this does not mean that 1st person is never specified for the feature *addressee* as some languages do mark the distinction between inclusive and exclusive 1st person in the plural (e.g. Arabic).

<sup>5</sup> Of course, in pro-drop languages pronouns can also be present (i.e. the feature content can be spelled out twice). It seems that the most natural way of looking at this is from the point of view of Relevance Theory (Sperber and Wilson 1995), whereby the effort involved in processing the same information twice must reward the hearer with extra cognitive effects. Indeed, it is well known that the use of overt pronouns in pro-drop languages is pragmatically highly marked (e.g. contrastiveness, topicalisation etc.).

<sup>6</sup> An alternative interpretation would be that the first person singular in Icelandic (and perhaps in German too) involves a phonologically empty suffix specified as

[+sp oad +sg], thus yielding the form *segi-Ø*). This seems like a rather straightforward assumption since this affix would correspond to a specific form (unlike, for example, in English) and therefore it would provide the contrast necessary for learnability conditions (Pinker 1984, see also chapter 1, section 2.2). The interesting question that arises on this analysis is why this null affix should favour the first person position.

<sup>7</sup> Duarte (1993, 1995) has shown that Brazilian Portuguese speakers occasionally drop pronouns in pro-drop contexts. However, the rate she reports (26%) is massively lower than the rates reported for pro-drop languages (these fluctuate between 60% and 85%; see for example Valian 1990, Vallduvi 1987, see also chapter 4, section 2.3). Moreover, Barbosa et al. (2005) have discussed evidence which suggests that subjectless clauses in Brazilian Portuguese and in pro-drop languages - which include earlier varieties of Brazilian Portuguese - are due to different licensing mechanisms.

<sup>8</sup> As far as Irish is concerned, I will use the conditional paradigm instead of the present indicative, since this is the only context in which pro-drop putatively occurs.

<sup>9</sup> Therefore, the feature specifications that appear in (17) are carried by elements other than affixes. This is unsurprising, since there seems to be no reason why Person and Number features should be exclusively associated with morphological agreement (cf. Koenenman 2000, see also section 2.1.7 below).

<sup>10</sup> Some proposals, notably Rohrbacher (1999), claim that the same agreement paradigms are responsible for both V-to-I movement and pro-drop. However, additional stipulations are necessary in order to account for the fact that some V-to-I languages do not allow pro-drop (e.g. Icelandic, Yiddish).

<sup>11</sup> I suspect that a similar intuition lies behind Rohrbacher's idea of "minimal distinctive marking", though our proposals differ radically in other respects.

<sup>12</sup> It has often been shown that - at the early stages of acquisition - children drop subject pronouns crosslinguistically, regardless of their target language (Radford 1990). Little can be said about the role of Cascade-updating during this stage. Presumably, this precedes the acquisition of phi-features (i.e. the children produce



either infinitives or unanalysed inflected forms), thus making any potential application of the universal rule in (3) totally vacuous.

<sup>13</sup> Note that the analysis that follows does not hinge on the –s morpheme carrying +sg in particular.

<sup>14</sup> Although the original data started from age 1;10, Andreu did not produce English VPs until age 3;2. As a result, the table in (10) does not include data prior to Andreu's third birthday.

<sup>15</sup> The monolingual studies are taken from Valian (1990) and involve two separate groups consisting of six and seven children respectively. The MLU and age values indicate mean MLU and mean age respectively (MLU ranging between 2.15 and 3.0 in one case and 3.08 and 3.94 in the other; age ranging between 2;3 and 2;8 in one case and 2;5 and 2;7 in the other).

<sup>16</sup> No specific value is given by the authors. However, they specify that at age 3;5 Andreu's MLU was 3.36, suggesting that – a month earlier – it must have been around 3.0.

<sup>17</sup> At MLU 1;5 Carlo had a rather low percentage of null-subjects in Italian (42%). However, as Serratrice et al point out, this happened to coincide with a particular interaction context which involved a large number of formulaic strings involving the demonstrative structure "questo è x" (this is x). Once these are counted out, Carlo's null-subject sentences raise to 86%.

<sup>18</sup> No information given by the authors.

<sup>19</sup> No file available for this stage.

<sup>20</sup> This study was concerned with the drop rate in relation to argument-type, not with the overtiness of subjects. The only data given by group (based on MLU) also contain object-drop and are therefore not fit for our comparison.

<sup>21</sup> But see Wijnen et al. 2001 and Aldridge et al. 1995 for some counterexamples.

<sup>22</sup> The data in (40) are taken from the following sources (in order of presentation): Italian: Guasti (1994), Shaeffer (1990); Spanish: Grinstead (1994); Catalan: Torrens (1992); French: Pierce (1992); English: Hoekstra and Hyamns (1998); Swedish: Platzak (1990); German Guasti (1994); Dutch Weverink (1989).

<sup>23</sup> In Hoekstra & Hyams the German data from Guasti (1994) are reported as follows:

(i)	German	S	2;1	46%
			2;2	40%

This seems to involve some error. Guasti (1994:11) reports that at age 2;1 S produced 53 infinitives and 46 finite verbs, which amounts to 53.5% RIs, not 46%.

<sup>24</sup> Evidence presented in Wexler (1998) suggests that RIs are also present in child Norwegian, though no detailed study has been discussed in the literature so no overall percentages are available.

<sup>25</sup> RI hypotheses which lie outside this dualism are also reported in the literature. Of these, I will briefly discuss Rizzi (1994) in section 3.3.5 below. I will ignore other proposals (e.g. Avrutin 1997, Ingham 1998, Phillips 1996), as their empirical inadequacy has been extensively discussed elsewhere in the literature (cf. Hoekstra and Hyams 1998, among others).

<sup>26</sup> For this ambiguity to be possible they are forced to assume that the clitic may not provide a sufficiently salient cue. However, to the extent that this assumption is legitimate, we must also grant ambiguity in the English case being discussed here since it follows the same line of reasoning (see Unsworth 2003 for a similar discussion). If anything, English root clauses present a better case for ambiguity since there is absolutely no phonological cue as to the finiteness of the verb, unlike the clitic cases where – though potentially weak – some cue is present.

<sup>27</sup> Though it is not obvious how this would work. In particular, if AgrS is set as *weak* by default (a position that Rizzi seems to endorse), transfer would then be impossible, as Cascade-updating is unable to override feature values (it only affects properties which are absent in one of the two languages).

## CONCLUSIONS

---

The aim of this thesis was to pursue an explanation of transfer effects in BFLA that is ultimately grounded in the principles underlying language acquisition. With this in mind, I presented a particular view of the language acquisition device, mostly based on standard linguistic theory. This involved the assumption that there is a Universal Grammar which aids children's linguistic development by providing them with the basic building blocks of human language from the initial state. I then suggested a number of properties which may be ascribed to the initial state. These included categorial features as well as features that encode semantic relations.

The main contribution made by this thesis has been the proposal of a model of acquisition that is based on the interaction between two systems: Minimal Paradigm and Cascade-updating. The first of these is a system dedicated to organising the developing lexicon into paradigms while the second is an informationally monotonic updating system whose role is to add newly acquired lexical information to those items that are not yet fully developed. The aim of this proposal was to provide an acquisitional model which is sufficiently conservative to avoid massive overshooting of the target and yet able to accommodate overgeneralisation as an inevitable consequence of the acquisition process.

A question was then raised in relation to the role of Cascade-updating in cases of bilingual acquisition. As recent findings on the organisation of the bilingual lexicon suggest that language separation need not involve separate lexicons, I concluded that the application of Minimal Paradigm may extend across the two languages. It was then suggested that the subsequent triggering of Cascade-updating would lead to overgeneralisation in bilinguals just as it does in monolinguals, thus resulting in the emergence of the phenomenon known as transfer effects.

A major part of the thesis was dedicated to assessing the empirical consequences of the Cascade-updating model. This involved considering various

sets of data from the literature, as well as discussing novel observations and presenting new experimental findings.

### **1. Findings**

In chapter 2 I discussed transfer effects within the domain of *wh*-elements from two separate case-studies. I then argued that these can be given a straightforward explanation within the Cascade-updating model since the lexical mappings at issue met the requirements necessary for Cascade-updating to apply. Data from two monolingual case-studies were also discussed. From this discussion it emerged that monolingual children display overgeneralisation effects in the exact same domain as bilinguals do and in relation to the very same feature, though different values happened to be involved. Although there has been speculation in the literature about transfer domains also being 'problematic' for monolinguals, our investigation revealed a much more intimate relation between the two phenomena. Not only is the *wh*-domain subject to transfer in bilinguals but it also feeds overgeneralisation in monolinguals. As the Cascade-updating model takes overgeneralisation and transfer to result from the same operation, these findings provide important evidence in its favour.

One of the case-studies was then analysed in more detail and with regard to other domains, namely modality and negation. In the case of modality, it was found that, given certain assumptions about the nature of modal relations, the transfer effects at hand could be easily accommodated within the Cascade-updating model. As to negation, the data at issue raised an important question in relation to the potentially different forms that transfer effects could take. In particular, we saw that one of the child's two languages, namely English, lacked Negative Quantifiers despite the fact that these are usually present in the speech of monolingual English children of the same age. Unlike transfer effects, this phenomenon is not immediately traceable to the child's other L1 as it does not involve the emergence of apparent literal translations. However, it was pointed out that this did not necessarily entail a different underlying process. It was then argued that this case of delay stemmed from the very same process that underlies transfer effects and that the absence of apparent literal translations was due to independent grammatical pressures.

The focus of chapter 3 was the Delay of Principle B Effects. This phenomenon has often attracted discussion in the bilingual literature as an example of a domain within which transfer effects are unattested. In particular, studies have investigated the idea of transfer effects between a language that completely lacks DPBE, namely

Greek, and a language that typically displays DPBE, such as English or Dutch. These studies have all reported the same findings, namely that DPBE is not subject to transfer. However, the very same studies have also maintained that the lack of DPBE in Greek is due to the fact that this language lacks 3<sup>rd</sup> person pronominals, the elements usually referred to as Greek pronouns being in fact demonstratives. It was then argued that although this categorial distinction was previously given little weight, it is in fact of major importance within a model that considers categorial features as a core notion in the triggering of transfer effects. If, as many have argued, the difference between demonstratives and personal pronouns involves categorial distinction, the grouping of these elements together is banned by Minimal Paradigm, and therefore incompatible with the Cascade-updating model. Consequently, once we assume the Cascade-updating model, the lexical make-up of pronominal elements in the languages concerned leads us to the conclusion that transfer of DPBE is necessarily ruled out.

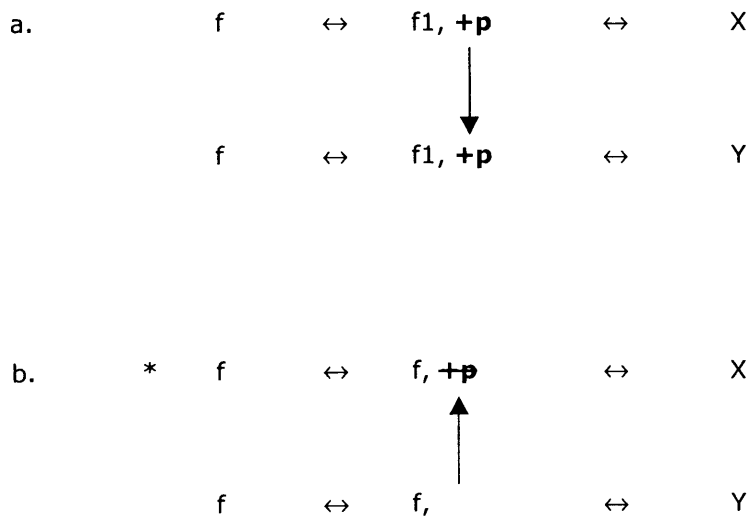
The discussion was then extended to monolingual cases, specifically in relation to languages that have both pronouns and clitics. It was pointed out that if we follow recent analyses of the syntactic properties of clitics and pronouns in assuming that these elements belong to different categories, the Cascade-updating model predicts that monolingual children should also be able to develop pronouns and clitics separately. Although clitics have often been reported to lack DPBE, children should fail to generalise this property over to pronouns. In other words, it should be possible for monolingual children to display DPBE in pronoun contexts even though they may not do so in clitic contexts. New experimental evidence was discussed which suggests that this is the case.

Finally, chapter 4 was concerned with a specific consequence of the Cascade-updating model, namely lexical dependency. As lexical mappings are the sole domain of application for Cascade-updating, typological properties which arise from the interaction of linguistic principles that are not lexically encoded must necessarily be immune to transfer effects. Two phenomena were discussed in relation to this: pro-drop and root infinitives, both of which were argued to arise from properties of the agreement system. As the notion of agreement is not encoded within any single lexical mapping, it was concluded that neither pro-drop nor RIs can be subject to Cascade-updating. A number of experimental findings were discussed which provide evidence for this claim.

## 2. Future Research

A claim that is central to the Cascade-updating model is informational monotonicity. Among other things, this property has the consequence of restricting the occurrence of transfer effects. In particular, it restricts directionality of transfer from items which possess a property to those which lack it, never the other way around. Suppose that two lexical items X and Y display a different syntactic behaviour in the adult language and that such difference is due to one particular property  $p$ . While X has the value  $[+p]$  Y has the value  $[-p]$ . During acquisition, there may be a stage in which X and Y will behave alike if neither of them has yet acquired a value for  $p$ . When, at a later stage, this property is set to  $[+p]$  for item X, the Cascade-updating model predicts that there is only one way in which the items can affect each other. While it is possible that Y is assigned the value  $[+p]$  through Cascade-updating, transfer from Y to X is ruled out. This is because such effect would involve transferring the lack of a value for  $p$ , an impossible outcome for an informationally monotonic system. In other words, given informational monotonicity, the process in (1a) is a logical possibility but that in (1b) is not:

1.



For any property  $p$ , if transfer of  $p$  occurs at all, it will be systematically from the item that has acquired a value to that which has not. In the hypothetical case described above the Cascade-updating model predicts that transfer effects can only occur in one direction, namely from the language that possesses the  $[+p]$  value to

the one that lacks a value for *p*. No other type of transfer is allowed in this circumstance.

This domain offers an important area of focus for further research. More language pairs need to be investigated in order to test whether linguistic behaviour associated with the lack of a property is potentially transferable.

Note that the above discussion on informational monotonicity has one further implication. It entails that, *modulo* paradigmatic restrictions, transfer effects may result if one of the two languages has acquired more information than the other with respect to one property. This has the consequence that although the direction of transfer effects will tend to be from the dominant to the less dominant language, it need not be so. It is entirely possible that the less dominant language has a developmental advantage with regard to a property, despite being disadvantaged in other respects. Indeed, it has sometimes been reported that some children seemed to transfer structures from the less dominant to the more dominant language (Genesee 1989, 1995). While rejecting the claim that language dominance is at the heart of transfer effects (cf. Volterra and Taeschner 1978), these findings are perfectly compatible with the Cascade-updating model. However, this area lacks systematic research and evidence of transfer from the less dominant to the more dominant language is very scant, if not anecdotal. More data needs to be collected and analysed in order to establish the likelihood of this state of affairs as well as its potential impact on the Cascade-updating model.

Finally, as briefly discussed in chapter 2, more research is needed with regard to the issue of unidirectionality of transfer. Some researchers (notably Müller 1998, and Hulk and Müller 2001) have suggested that *directionality* of transfer is dependent on the languages at issue. For example, Müller (1998) argues that in the acquisition of relative clauses in German/French bilinguals transfer is predicted to occur systematically from French to German. Within the Cascade-updating model, however, unidirectionality of transfer need not be dependent on the language pair. In cases where the two target languages have a conflicting value for a given property (i.e. [+P] v. [-P]) the direction of transfer will depend on which value is acquired first. While child A may display transfer effects from the French to the German paradigm, it is entirely possible that child B acquiring the same language pair could well display exactly the same transfer in the opposite direction, *modulo* universal restrictions on the properties at issue. In other words, directionality of transfer need not be constant across bilinguals. Future research should undoubtedly consider the importance of testing this possibility.

## References

- Abney, S.P. 1987. The English Noun Phrase and its Sentential Aspect. Ph.D dissertation, MIT Cambridge.
- Ackema, P. 2002. A morphological approach to the absence of expletive PRO. *UCL Working Papers in Linguistics* 14: 291-319.
- Ackema, P. and A. Neeleman. 2003. Context-sensitive spell-out. *Natural Language and Linguistic Theory* 21: 681-735.
- Acquaviva, P. 1999. Negation and operator dependencies. *Lingua* 108: 137-174.
- Adger, D. 2003. *Core Syntax. A Minimalist Approach*. Oxford University Press, Oxford.
- Aldridge, M., R Borsley, and S. Clack. 1995. The Acquisition of Welsh Clause Structure. *Proceedings of the Boston University Conference on Language Development*. (BUCLD 19). Cascadilla Press, Somerville MA.
- Alexiadou, A. and E. Anagnostopoulou. 1998. Parametrizing Agr. Word Order, V-Movement and EPP-checking. *Natural Language and Linguistic Theory* 16:491-539.
- Allen, S. E. M. and Schröder, H. 2003. Preferred argument structure in early Inuktitut spontaneous speech data. In J. W. Du Bois, L. E. Kumpf and W. J. Ashby (eds.), *Preferred Argument Structure: grammar as architecture for function*. John Benjamins, Amsterdam.
- Andrews, A. 1990. Unification and morphological blocking. *Natural Language and Linguistic Theory* 8: 507-557.
- Antelmi, D. 1992. L' Ipotesi Maturazionale Nell' Acquisizione Del Linguaggio: Indagine Longitudinale su una Bambina Italiana. Università degli Studi di Padova.
- Antelmi, D. 1997. *La prima grammatica dell'italiano : indagine longitudinale sull'acquisizione della morfosintassi italiana*. Il Mulino, Bologna.
- Aronoff, M. 1976. *Word Formation in Generative Grammar*. MIT Press, Cambridge, Mass.
- Avrutin, S. 1997. Events as units of discourse representation in Root Infinitives. *MIT Working Papers in Linguistics* 12: 65-90
- Avrutin, S. and K. Wexler 1992. Development of Principle B in Russian: coindexation at LF and coreference. *Language Acquisition* 4: 259-306.
- Baauw, S. 2002. *Grammatical Features and the Acquisition of Reference*. Routledge, London.
- Baauw, S., L. Escobar and W. Philip. 1997. A delay of Principle B effect in Spanish speaking children. In: A. Sorace, C. Heycock and R. Shillcock (eds.), *Proceedings of the GALA '97 Conference on Language Acquisition*. HCRC Edinburgh.



- Baker, C. L. 1979. Syntactic theory and the projection problem. *Linguistic Inquiry* 10: 533-581.
- Barbosa, P., M. E. Duarte and M. Kato (2005) 'Null Subjects in European and Brazilian Portuguese'. *Journal of Portuguese Linguistics* 4:11-52. Lisboa: Edições Colibri.
- Belletti, A. 1990. *Generalized Verb Movement*. Rosenberg and Sellier, Turin.
- Belletti, A. 1999. Italian/Romance clitics: structure and derivation. In Henk van Riemsdijk (ed.) *Clitics in the Languages of Europe*. Mouton de Gruyter, Berlin.
- Bellugi, U. 1967. The acquisition of the system of negation in children's speech. Ph.D. dissertation, Harvard University.
- Berger-Morales, J. and M. Salustri. 2003. Supporting the separate systems hypothesis: a case study in bilingual acquisition of Italian and German. In A.T. Perez-Leroux & Y. Roberge (eds.), *Romance Linguistics: Theory and Acquisition*. John Benjamins, Amsterdam.
- Berger-Morales, J., M. Salustri and J. Gilkerson. 2003. Root Infinitives in the spontaneous speech of two bilingual children: evidence for separate grammatical systems. Paper presented at the *4th International Symposium on Bilingualism*. Arizona State University.
- Berko, J. 1958. The child's learning of English morphology. *Word* 14:150-177.
- Berndt, R. 1956. *Form und Funktion des Verbums im nördlichen Spätmittelenglischen*. Max Niemeyer, Halle.
- Bhat, D. N. S. 2004. *Pronouns*. Oxford University Press, Oxford.
- Bickerton, D. 1990. *Language and Species*. Chicago University Press, Chicago.
- Blevins, J. 1995. Syncretism and paradigmatic opposition. *Linguistics and Philosophy* 18: 113-152.
- Bloom, P. 1993. Grammatical continuity in language development: The case of subjectless sentences. *Linguistic Inquiry* 24:721-734.
- Bloom, L., P. Lightbown, and L. Hood. 1975. Structure and variation in child language. *Monographs of the Society for Research in Child Development* 40.
- Borer, H. 1984. *Parametric Syntax*. Foris Publications, Dordrecht.
- Borer, H. 1986. I-Subjects. *Linguistic Inquiry* 17: 375-416.
- Borer, H. and K. Wexler. 1987. The Maturation of Syntax. In: T. Roeper and E. Williams (eds) *Parameter Setting*. Reidel, Dordrecht.
- Bouma, L. 1975. On contrasting the semantics of the modal auxiliaries of German and English. *Lingua* 37: 313-39.

- Bowerman, M. 1982. Evaluating competing linguistic models with language acquisition data: implications of developmental errors with causative verbs. *Quaderni di Semantica* 3: 5-66.
- Brody, M. 2002. Readings in Syntax. Lecture series delivered in the Spring 2002, University College London.
- Brown, R. 1973. *A First Language: The Early Stages*. Penguin, Harmondsworth.
- Brown, S. 1999. *The Syntax of Negation in Russian*. CSLI Publications, Stanford.
- Brugè, L. 2002. The Position of Demonstratives in the Extended Nominal Projection. In Guglielmo Cinque (eds.) *Functional Structure in DP and IP*. Oxford University Press, Oxford.
- Büring, D. 2005. *Binding Theory*. Cambridge University Press, Cambridge.
- Campbell, A. 1959. *Old English Grammar*. Oxford University Press, Oxford.
- Campbell, R. 1996. Specificity Operators in SpecDP. *Studia Linguistica* 50, 161-188.
- Caramazza, A. and I. Brones. 1980. Semantic classification by bilinguals. *Canadian Journal of Psychology* 34: 77-81.
- Cardinaletti, A. and M. Starke. 1999a. The typology of structural deficiency: on the three grammatical classes. In Henk van Riemsdijk (ed.) *Clitics in the languages of Europe*. 145-233. Mouton de Gruyter, Berlin.
- Cardinaletti, A. and Starke, M. 1999b. Responses and demonstratives. In Henk van Riemsdijk (ed.) *Clitics in the languages of Europe*. Mouton de Gruyter, Berlin.
- Carstairs, A. 1983. Paradigm economy. *Journal of Linguistics* 19: 115-128.
- Castro, D. and E. Gavrusseva. 2003. Finiteness and aspect in Spanish/English bilingual acquisition. *First Language* 23: 171-192.
- Chee, M. W., L. Tan, and E. W. L. Thiel. 1999. Mandarin and English single word processing studied with functional magnetic resonance imaging. *The Journal of Neuroscience* 19: 3050-3056.
- Chen, H. C. and M. L. Ng. 1989. Semantic facilitation and translation priming effects in Chinese-English bilinguals. *Memory and Cognition* 17: 454-462.
- Chien, Y. and K. Wexler. 1990. Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics. *Language Acquisition* 1: 225-295.
- Chomsky, N. 1962. A transformational approach to syntax. In A.A. Hill (ed.) *Proceedings of the Third Texas Conference on Problems of Linguistic Analysis in English*. University of Texas Press, Austin.
- Chomsky, N. 1965. *Aspects of the Theory of Syntax*. MIT Press, Cambridge, Mass.
- Chomsky, N. 1973. Conditions on transformations. in: S.R. Anderson & P. Kiparsky, *A festschrift for Morris Halle*. Holt, Rinehart and Winston, New York.

- Chomsky, N. 1977. *Essays on Form and Interpretation*. Elsevier North-Holland, Amsterdam.
- Chomsky, N. 1981. *Lectures on Government and Binding*. Foris Publications, Dordrecht.
- Chomsky, N. 1982. *Some Concepts and Consequences of the Theory of Government and Binding*. MIT Press, Cambridge, Mass.
- Chomsky, N. 1986a. *Barriers*. MIT Press, Cambridge, Mass.
- Chomsky, N. 1986b. *Knowledge of Language: Its Nature, Origin, and Use*. Praeger, New York.
- Chomsky, N. 1993. A minimalist program for linguistic theory. In K. Hale and S. Keyser (eds.), *The View from Building 20*. MIT Press, Cambridge, Mass.
- Chomsky, N. 1995. *The Minimalist Program*. MIT Press, Cambridge, Mass.
- Chomsky, N. 2000. *New Horizons in the Study of Language and Mind*. Cambridge University Press, Cambridge.
- Chomsky, N. and M. Halle. 1968. *The Sound Pattern of English*. MIT Press, Cambridge, Mass.
- Cipriani, P. et al. 1989. *Protocolli Diagnostici e Terapeutici nello Sviluppo e nella Patologia del Linguaggio (1/84)* Italian Ministry of Health: Stella Maris Foundation.
- Cipriani, P., A. M. Chilosi, P. Bottari and L. Pfanner. 1991. *L'Acquisizione della Morfosintassi: Fasi e Processi*. Unipress, Padua.
- Cormack, A. and N. Smith. 2000. Head movement and negation in English. *Transactions of the Philological Society* 98: 49-85.
- Crain, S. and R. Thornton. 1998. *Investigations in Universal Grammar. A Guide to Experiments on the Acquisition of Syntax and Semantics*. MIT Press, Cambridge, Mass.
- De Groot, A. M. B. 1993. Word-type effects in bilingual processing tasks: support for a mixed-representational system. In Schreuder, R. and B. Weltens (ed.) *The Bilingual Lexicon*. Benjamins, Amsterdam.
- De Groot, A. M. B., L. Dannenburg, and J. G. Van Hell. 1994. Forward and backward word translation by bilinguals. *Journal of Memory and Language* 33: 600-629.
- De Houwer, A. 1990. *The Acquisition of Two Languages from Birth: A Case Study*. Cambridge University Press, Cambridge.
- de Swart, H. 1998. Licensing of negative polarity items under inverse scope. *Lingua* 105: 175-200.
- Dechaine, R. M. and M. Wiltschko. 2002. Decomposing pronouns. *Linguistic Inquiry* 33: 409-442.

- Delfitto, D. and N. Corver. 1993. Feature asymmetry and the nature of pronoun movement. *OTS Working Papers*, Utrecht University.
- Demuth, K. 1994. On the Underspecification of functional categories in early grammars. In B. Lust, M. Suñer, and J. Whitman (eds.), *Syntactic Theory and First Language Acquisition*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Déprez, V. 1994. Underspecification and functional projections. In B. Lust, M. Suñer, and J. Whitman (eds.), *Syntactic Theory and First Language Acquisition*. Lawrence Erlbaum Associates, Hillsdale, New Jersey.
- Deuchar, M. and S. Quay. 2000. *Bilingual Acquisition: Theoretical Implications of a Case Study*. Oxford University Press, Oxford.
- Diesing, M. 1992. *Indefinites*. MIT Press, Cambridge, Mass.
- Diesing, M. 1999. Comments on Cardinaletti and Starke "The typology of structural deficiency". In Henk van Riemsdijk (ed.) *Clitics in the languages of Europe*. 145-233. Mouton de Gruyter, Berlin.
- Diessel, H. 1999. *Demonstratives: Form Function and Grammaticalization*. Benjamins, Amsterdam.
- Dijkstra, T. and W. J. B. van Heuven. 2002. The architecture of the bilingual word recognition system: from identification to decision. *Bilingualism: Language and Cognition* 5: 175-197.
- Döpke, S. 2000. *Cross-linguistic Structures in Simultaneous Bilingualism*. Benjamins, Amsterdam.
- Doron, E. 1988. On the complementarity of subject and subject-verb agreement. In M. Barlow and C. Fergusson (eds.), *Agreement in Natural Language: Approaches, Theory, Description*. CSLI Publications, Stanford.
- Dresher, B. E and J. D. Kaye. 1990. A computational learning model for metrical phonology. *Cognition* 34: 137-195.
- Dressler, W. 1985. On the predictiveness of natural morphology. *Journal of Linguistics* 21: 321-338.
- Duarte, M. E. L. 1993. Do pronome nulo ao pronome pleno: a trajetória do sujeito no português do Brasil. In I. Roberts and M. A. Kato (eds.) *Português Brasileiro: uma viagem diacrônica*. Campinas, SP, Ed. da Unicamp.
- Duarte, M. E. L. 1995. *A perda do princípio Evite Pronome no Português Brasileiro*. PhD dissertation, Universidade de Campinas.
- Ehrman, M. E. 1966. *The Meanings of the Modals in Present-day English*. The Hague, Mouton.

- Elliott, W. N., and K. Wexler. 1985. A principled theory of categorical acquisition. Paper presented at the *Meeting of the Northeastern Linguistic Society*, XVI, Montreal.
- Escobar, L. and A. Gavarro. 1999. *The Acquisition of Catalan Clitics and its Implications for Complex Verb Structure*. Report de Recerca, Universitat Autònoma de Barcelona.
- Eisenbeiss, S. 2000. The acquisition of the Determiner Phrase in German child language. In Friedemann, M.-A. and L. Rizzi (eds.) *The Acquisition of Syntax: Studies in Comparative Developmental Linguistics*, 26-62. Longman Linguistics Library Series, Geneva.
- Espinal, M. T. 2000. On the semantic status of n-words in Catalan and Spanish. *Lingua* 110: 557-580.
- Etxepare, R. and K. K. Grohmann. 2005. Towards a grammar of adult Root Infinitives. *Proceedings of West Coast Conference in Formal Linguistics* 14. 24: 129-137. CSLI, Stanford University.
- Fodor, J. D. and W. G. Sakas. 2005. The Subset Principle in syntax: costs of compliance. *Journal of Linguistics* 41: 513-569.
- Fortescue, M. 1984. *West Greenlandic*. Croom Helm, London.
- Fox, D. 2000. *Economy and Semantic Interpretation*. MIT Press, Cambridge, Mass.
- Franks, S. and C. Rudin. 2004. Bulgarian clitics as K0 heads. Paper presented at *Formal Approaches to Slavic Linguistics* 13, University of South Carolina.
- French, R. M. 1998. A simple recurrent network model of bilingual memory. *Proceedings of the Annual Cognitive Science Society Conference*. 20: 368-373.
- French, R. M. and M. Jacquet. 2004. Understanding bilingual memory: models and data. *Trends in Cognitive Sciences* 8: 87-93.
- Gaume, B., K. Duvignau, O. Gasquet, and M.-D. Gineste. 2002. Forms of meaning, meaning of forms. *Journal of Experimental and Theoretical Artificial Intelligence* 14: 61-74.
- Gelderen, E. van 2000. *A History of English Reflexive Pronouns: Person, Self, and Interpretability*. Benjamins, Amsterdam.
- Genesee, F. 1989. Early bilingual development: one language or two? *Journal of Child Language* 6: 191-179.
- Genesee, F. 1995. *Integrating Language and Content: Lessons from Immersion*. National Center for Research on Cultural Diversity and Second Language Learning. Washington, DC.
- Genesee, F., E. Nicoladis, and J. Paradis. 1995. Language differentiation in early bilingual development. *Journal of Child Language* 22: 611-631.

- Gerken, L. 1994. Young children's representation of prosodic phonology. *Journal of Memory and Language* 33: 19-38.
- Gerlach, B. 2002. *Clitics between Syntax and the Lexicon*. Benjamins, Amsterdam.
- Giannakidou, A. 1998. *Polarity Sensitivity as (Non)Veridical Dependency*. Benjamins, Amsterdam.
- Giannakidou, A. 2000. Negative ...Concord? *Natural Language and Linguistic Theory* 18:457-523.
- Giannakidou, A. 2006. N-words and Negative Concord. In M. Everaert and H. van Riemsdijk (eds.), *The Blackwell Companion to Syntax*. Blackwell, Oxford.
- Giorgi, A. and F. Pianesi. 1997. *Tense and Aspect: from Semantics to Morphosyntax*. Oxford University Press, Oxford.
- Giusti, G. 1993. *La Sintassi dei Determinanti*. Unipress, Padua.
- Giusti, G. 2002. The functional structure of noun phrases: a bare phrase structure approach. In Guglielmo Cinque (eds.) *Functional Structure in DP and IP*. Oxford University Press, Oxford.
- Gold, E. M. 1967. Language identification in the limit. *Information and Control* 10: 447-474.
- Grinstead, J. 1994. Consequences of the maturation of number morphology in Spanish and Catalan. M.A. Thesis, UCLA.
- Grodzinsky, J. and T. Reinhart. 1993. The innateness of binding and coreference. *Linguistic Inquiry* 24: 69-102.
- Grodzinsky, Y., K. Wexler, Y. Chien and S. Marakovitz. 1989. The break-down of Binding. Paper presented at the *Academy of Aphasia*, Santa Fe, New Mexico,
- Guausti, M.-T. 1994. Verb syntax in Italian child grammar: finite and non-finite verbs. *Language Acquisition* 3: 1-40.
- Haegeman, L. and R. Zanuttini. 1991. *Negative Heads and the Neg Criterion*. Oxford University Press, Oxford.
- Hale, K. and S. J. Keyser. 1993. On argument structure and the lexical expression of syntactic relations. In K. Hale and S. Keyser (eds.), *The View from Building 20*. MIT Press, Cambridge, Mass.
- Hale, M. and C. Reiss. 2003. The Subset Principle in phonology: why the tabula can't be rasa. *Journal of Linguistics* 39:219-244.
- Heim, I. 1982. The Semantics of Definite and Indefinite Noun Phrases. PhD dissertation, University of Massachusetts at Amherst.
- Higginbotham, J. and R. May. 1981. Questions, quantifiers, and crossing. *The Linguistic Review* 1: 41-79.

- Hoekstra, T. and N. Hyams. 1995. The syntax and interpretation of dropped categories in child language: a unified account. *Proceedings of West Coast Conference in Formal Linguistics* 14. CSLI, Stanford University.
- Hoekstra, T. and N. Hyams. 1998. Aspects of Root Infinitives. *Lingua* 106: 81-112.
- Hoekstra, T. and P. Jordens. 1994. From adjunct to head. In T. Hoekstra and B. Schwartz (eds.), *Language Acquisition Studies in Generative Grammar*. Benjamins, Amsterdam.
- Holmberg, A. 2003. Null subjects and uninterpretable features: evidence from Finnish. *Durham Working Papers in Linguistics* 9: 11-24.
- Hudson, R. 2007. *Language Networks: The new Word Grammar*. Oxford University Press, Oxford.
- Hulk, A. and N. Müller. 2000. Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition* 3: 227-244.
- Hulk, A. and N. Müller. 2001. Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition* 4: 1-21
- Hyams, N. 1996. The underspecification of functional categories in early grammar. In H. Clahsen (ed.), *Generative Perspectives on Language Acquisition*, John Benjamins, Amsterdam.
- Illes, J. et al. 1999. Convergent cortical representation of semantic processing in bilinguals. *Brain and Language* 70: 347-363.
- Ingham, R. 1988. Tense without agreement in early clause structure. *Language Acquisition* 7: 51-81.
- Ingham, R. 1992. The optional subject phenomenon in young children's English: a case study. *Journal of Child Language* 19: 133-51.
- Ingram, D. 1974. Phonological rules in young children. *Journal of Child Language* 1: 49-64.
- Jackendoff, R. 1975. Morphological and semantic regularities in the lexicon. *Language* 51: 639-671.
- Jackendoff, R. 1987. The status of thematic relations in linguistic theory. *Linguistic Inquiry* 17: 369-411.
- Jackendoff, R. 1990. *Semantic Structures*. MIT Press, Cambridge, Mass.
- Jackendoff, R. 1997. *The Architecture of the Language Faculty*. MIT Press, Cambridge, Mass.
- Jaeggli, O. 1982. *Topics in Romance Syntax*. Foris Publications, Dordrecht.

- Jaeggli, O. and K. Safir. 1989. The null-subject parameter and parameter theory. In O. Jaeggli and K. Safir (eds.), *The Null Subject Parameter*. Kluwer Academic Publishers, Dordrecht.
- Jakubowicz, C. 1993. Linguistic theory and language acquisition facts: reformulation, maturation, or invariance of Binding Principles. In E. Reuland and W. Abraham (eds.), *Knowledge and Language*, Vol. 1, Kluwer Academic Publishers, Dordrecht.
- Joos, M. 1964. *The English Verb: Form and Meanings*. The University of Wisconsin, Madison and Milwaukee.
- Juan-Garau, M. and C. Perez-Vidal. 2000. Subject realization in the syntactic development of a bilingual child. *Bilingualism: Language and Cognition* 173-191.
- Kahane, S. and A. Polguère. 2001. Formal foundation of lexical functions. *Actes du colloque COLLOCATION*, 8-15. Toulouse.
- Karttunen, L. 1984. Features and values. *Proceedings of the Tenth International Conference on Computational Linguistics*. Stanford, California.
- Kayne, R. 1975. *French Syntax: The Transformational Cycle*. MIT Press, Cambridge, Mass.
- Kayne, R. 1983. Chains, categories external to S, and French Complex Inversion. *Natural Language and Linguistic Theory* 1: 107-139.
- Kayne, R. 1989. Facets of Romance past participle agreement. in P. Beninca, ed., *Dialect Variation and the Theory of Grammar*. Foris Publications, Dordrecht.
- Kayne, R. S. 1991. Romance clitics, verb movement, and PRO. *Linguistic Inquiry* 22: 647:686.
- Keatley, C. W., et al. 1994. Asymmetrical cross-language priming effects. *Memory and Cognition* 22: 70-84.
- Kiparsky, P. 1973. 'Elsewhere' in phonology. In S. Anderson and P. Kiparsky (eds.), *A Festschrift for Morris Halle*. Holt, Rinehart and Winston, New York.
- Klein, D., Milner B., et al. 1995. The neural substrates underlying word generation: a bilingual function-imaging study. *Proceedings of the National Academy of Sciences*. 92: 2899-2903.
- Klima, E. S. and U. Bellugi. 1966. Syntactic regularities in the speech of children. *Proceedings of the 1966 Edinburgh Conference*. Edinburgh University Press, 183-208. Edinburgh.
- Klinge, A. 1993. The English modal auxiliaries: from lexical semantics to utterance interpretation. *Journal of Linguistics* 29: 315-357.
- Koeneman, O. 2000. The Flexible Nature of Verb Movement. Ph.D diss. Utrecht University.



- Koenenman, O. and A. Neeleman. 2001. Predication, verb movement and the distribution of expletives. *Lingua* 111: 189-233.
- Kratzer, A. 1977. What 'must' and 'can' must and can mean. *Linguistics and Philosophy* 1:337-355.
- Kwan-Terry A. 1986. The acquisition of word-order in English and Cantonese interrogative sentences: a Singapore case study. *RELJ Journal* 17: 14-39.
- Labov, W. and T. Labov 1978. Learning the syntax of Questions. In R. Campbell and P. Smith (eds.). *Recent Advances in the Psychology of Language III*, 1-44. Plenum Press, New York.
- Ladusaw, W. A. 1992. Expressing negation. In C. Baker and D. Dowty (Eds.) *Proceedings of the Second Conference on Semantic and Linguistic Theory*. The Ohio State University.
- Ladusaw, W. A. 1994. Thetic and categorial, stage and individual, weak and strong. Talk presented at the *Conference on Semantic and Linguistic Theory (SALT)*. University of California, Santa Cruz.
- Laka, I. 1990. Negation in Syntax: On the Nature of Functional Categories and Projections. Ph.D. dissertation, MIT Cambridge.
- Lakatos, I. 1970. Falsification and the methodology of scientific research programs. In I. Lakatos and A. Musgrave (eds.) *Criticism and the Growth of Knowledge*. Cambridge University Press, Cambridge.
- Lapointe, S. 1980. A theory of grammatical agreement. Ph.D. dissertation, University of Massachusetts, Amherst.
- Legate, J. A. 1999. The Morphosyntax of Irish Agreement. Ms., MIT.
- Leiss, E. 1997. Synkretismus und Natürlichkeit. *Folia Linguistica* 31: 133-160.
- Li, P. and I. Farkas. 2002. A self-organizing connectionist model of bilingual processing. In R. H. J. Altarriba (eds.) *Bilingual Sentence Processing*. Elsevier.
- Linebarger, M. 1987. Negative polarity and grammatical representation. *Linguistics and Philosophy* 10, 325-387.
- Logan, H. M. 1973. *The Dialect of the Life of Saint Katherine*. The Hague, Mouton.
- Luger, G. and W. Stubblefield. 1993. *Artificial Intelligence. Structures and Strategies for Complex Problem Solving*, 2nd edition. Benjamin-Cummings, New York.
- Lyons, J. 1968. *Introduction to Theoretical Linguistics*. Cambridge University Press, Cambridge.
- Lyons, J. 1977. *Semantics*. Cambridge University Press, Cambridge.

- MacWhinney, B. 1978. Processing a first language: the acquisition of morphophonology. *Monographs of the Society for Research in Child Development* 43.
- MacWhinney, B. 2000. *The CHILDES Database: Tools for Analyzing Talk*. 3<sup>rd</sup> Edition. Volume 2: The Database. Lawrence Erlbaum Associates, Mahwah, NJ.
- MacWhinney, B. and C. Snow. 1985. The Child Language Data Exchange System. *Journal of Child Language* 12: 271-296.
- MacWhinney, B. and C. Snow. 1990. The Child Language Data Exchange System: an update. *Journal of Child Language* 17:457-472.
- Mathieu, E. 2001. On the nature of French N-words. *UCL Working Papers in Linguistics* 13: 319-352.
- May, R. 1977. The grammar of quantification. Ph.D. dissertation, MIT Cambridge.
- May, R. 1985. *Logical Form in Natural Language*. MIT Press, Cambridge, Mass.
- May, R. 1989. Interpreting Logical Form. *Linguistics and Philosophy* 12: 387-435.
- McCloskey, J. and K. Hale. 1984. On the syntax of person-number inflection in Modern Irish. *Natural Language and Linguistic Theory* 1: 487-533.
- McKee, C. 1992. A comparison of pronouns and anaphors in Italian and English acquisition. *Language Acquisition* 2: 21-54.
- Meisel, J. 1989. Early differentiation of language in bilingual children. In Hyltenstam, K. and L. Obler (eds.) *Bilingualism Across the Lifespan*. Cambridge University Press, Cambridge.
- Meisel, J. 1994. Getting FAT. In J. Meisel (ed.), *Bilingual First Language Acquisition: French and German Grammatical Development*. Benjamins, Amsterdam.
- Montague, R. 1973. The proper treatment of quantification in ordinary English. In J. Hintikka, J. Moravcsik & P. Suppes (Eds.), *Approaches to Natural Language*. Reidel, Dordrecht.
- Müller, G. 2005. Pro-Drop and Impoverishment. Ms. Universität Leipzig.
- Müller, N. 1998. Transfer in bilingual language acquisition. *Bilingualism: Language and Cognition* 1: 151-171.
- Müller, N., K. Cantone, T. Cupisch, and K. Schmitz. 2002. About the linguistic influence in bilingual first language acquisition: Italian-German. *Linguistische Berichte* 190: 157-206.
- Neeleman, A. and F. Weerman. 1999. *Flexible Syntax*. Kluwer Academic Publishers, Dordrecht.
- Neeleman, A. and H. van de Koot. 2002. Bare resultatives. *Journal of Comparative Germanic Linguistics* 6: 1-52.

- Neeleman, A. and K. Szendrői. 2005. Radical Pro-Drop and the Morphology of Pronouns. Ms. UCL/Utrecht University.
- Noguchi, T. 1997. Two types of pronouns and variable binding. *Language* 73: 770-797.
- Noyer, R. 1992. Features, positions and affixes in autonomous morphological structure. Ph.D. dissertation, MIT Cambridge.
- Padilla, J. 1990. *On the Definition of Binding Domains in Spanish*. Kluwer Academic Publishers, Dordrecht.
- Palmer, F. R. 1990. *Modality and the English Modals*. Longman, London and New York.
- Palmer, F. R. 2001. *Mood and Modality*. Cambridge University Press, Cambridge.
- Panagiotidis, P. 2000. Demonstratives, determiners and operators: the case of Greek. *Lingua* 110:717-742.
- Papafragou, A. 1998. The acquisition of modality: implications for theories of semantic representation. *Mind and Language* 13: 370-399
- Papangeli, D. 2000. Clitic doubling in Modern Greek: a head- complement relation. *UCL Working Papers in Linguistics*
- Paradis, J. and F. Genesee. 1996. Syntactic acquisition in bilingual children. *Studies in Second Language Acquisition* 18: 1-25.
- Peng, L. L. 1998. The development of wh-questions in a Cantonese/English bilingual child. Unpublished M.Phil. Thesis, Chinese University of Hong Kong
- Perkins, M. R. 1982. The core meaning of the English modals. *Journal of Linguistics* 18: 245-73.
- Perlmutter, D. 1971. *Deep and Surface Constraints in Generative Grammar*. Holt, Rinehart and Winston, New York.
- Perovic, A. 2004. Knowledge of Binding in Down Syndrome: Evidence from English and Serbo-Croatian. Ph.D dissertation, University College London.
- Pesetsky, D. 1997. OT and syntax: movement and pronunciation. In D. Arcangeli and D. T. Langendoen (eds.) *Optimality Theory: an Overview*. Blackwell, Oxford.
- Pesetsky, D. 2000. *Phrasal Movement and Its Kin*. MIT Press, Cambridge, Mass.
- Philip, W. and P. Coopmans. 1996. The double Dutch delay of Principle B effect. *Proceedings of the Boston University Conference on Language Development*. 576-587. Cascadilla Press, Somerville MA.
- Phillips, C. 1996. Root Infinitives are finite. *Proceedings of the Boston University Conference on Language Development*. (BUCLD 20). Cascadilla Press, Somerville MA.

- Pierce, A. 1992. *Language Acquisition and Syntactic Theory: A Comparative Analysis of French and English Child Grammars*. Kluwer Academic Publishers, Dordrecht.
- Pinker, S. 1984. *Language Learnability and Language Development*. Harvard University Press, Cambridge, Ma.
- Pinker, S. 1989. *Learnability and Cognition: the acquisition of argument structure*. MIT Press, Cambridge, Mass.
- Platzack, C. 1990. A grammar without functional categories: a syntactic study of early Swedish child language. *Nordic Journal of Linguistics* 13: 107-126.
- Platzack, C. 1999. The vulnerable C-domain. Paper presented at the 4th GALA Conference, Potsdam.
- Platzack, C. and A. Holmberg. 1989. The role of AGR and finiteness in Germanic VO languages. *Working papers in Scandinavian syntax* 43: 51-76.
- Poeppel, D. and K. Wexler. 1993. The full competence hypothesis of clause structure in early German. *Language* 69: 1-33.
- Pogatscher, A. 1901. Unausgedrucktes Subject im Altenglischen. *Anglia*. 23: 261-301.
- Poletto, C. and P. Benincà. 2004. Topic, Focus and V2: defining the CP sublayers. In L. Rizzi (ed.) *The Structure of IP and CP. The Cartography of Syntactic Structures*, vol. 2, Oxford University Press, New York.
- Pollock, J.-Y. 1989. Verb movement, Universal Grammar, and the structure of IP. *Linguistic Inquiry* 20: 365-424.
- Poole, I. 1934. Genetic development of articulation of consonants sounds in speech. *Elementary English Review* 2: 159-61.
- Postal, P. M. 1969. On so-called 'pronouns' in English. In D.A. Reibel and S.A. Schane (eds.) *Modern Studies in English: Readings in Transformational Grammar*. Englewood Cliffs, Prentice-Hall.
- Pranka, P. 1983. Syntax and Word Formation. Ph.D. dissertation, MIT Cambridge.
- Radford, A. 1988. *Transformational Grammar*. Cambridge University Press, Cambridge.
- Radford, A. 1990. *Syntactic Theory and the Acquisition of English Syntax*. Blackwell, Oxford.
- Radford, A. 1995. Children - architects or brickies? *Proceedings of the Boston University Conference on Language Development*. (BUCLD 19). Cascadilla Press, Somerville MA.

- 
- Radford, A. 1996. Towards a structure-building model of acquisition. In H. Clahsen (ed.), *Generative Perspectives on Language Acquisition*, John Benjamins, Amsterdam.
- Reinhart, T. 1986. Center and periphery in the grammar of anaphora. In Lust et al (eds.) *Studies in the Acquisition of Anaphora*. Reidel, Dordrecht.
- Reinhart, T. 1997. Quantifier scope: how labor is divided between QR and choice functions. *Linguistics and Philosophy* 20: 335-397.
- Reinhart, T. 2000. Strategies of anaphora resolution. In I. Bennis, M. Everaert, and E. Reuland (ed.), *Interface Strategies*. Royal Academy of Arts and Sciences, Amsterdam.
- Rizzi, L. 1982. *Issues in Italian Syntax*. Foris Publications, Dordrecht.
- Rizzi, L. 1986. On chain formation. In: H. Borer (ed.) *Syntax and Semantics: Vol. 19 The Syntax of Pronominal Clitics*. Harcourt Brace Javanovich Publishers, London.
- Rizzi, L. 1990. Speculations on verb second. In J. Mascaro and M. Nespor (eds.) *Grammar in Progress*. Foris, Dordrecht.
- Rizzi, L. 1993/1994. Some notes on linguistic theory and language development: the case of Root Infinitives. *Language Acquisition* 3: 371-393.
- Rizzi, L. 1994. Early null subjects and root null subjects. In T. Hoekstra and B. Schwartz (eds.), *Language Acquisition Studies in Generative Grammar*. Benjamins, Amsterdam.
- Rizzi, L. 1996. Residual verb second and the wh-criterion. In A. Belletti and L. Rizzi (eds.) *Parameters and Functional Heads*. Oxford University Press, Oxford.
- Rizzi, L. 1997. The fine structure of the left periphery. In L. Haegeman (ed.) *Elements of Grammar*. Kluwer, Dordrecht.
- Rizzi, L. 2001. On the position "Int(errogative)" in the left periphery of the clause. In G. Cinque and G. Salvi (eds.) *Current Studies in Italian Syntax. Essays Offered to Lorenzo Renzi*. Elsevier, Amsterdam.
- Roberts, I. 1993. *Verbs and Diachronic Syntax*. Kluwer Academic Publishers, Dordrecht.
- Roberts, I. 2003. Some Consequences of a Deletion Analysis of Null Subjects. Ms. Downing College, University of Cambridge.
- Roeper, T. and J. Weissenborn. 1990. How to make parameters work: comments on Valian. In L. Frazier and J. de Villiers (eds) *Language Processing and Language Acquisition*. Kluwer, Dordrecht.
- Rohrbacher, B. 1994. The Germanic VO languages and the full paradigm: a theory of V to I raising. Ph.D. dissertation, University of Massachusetts, Amherst.

- Rohrbacher, B. 1999. *Morphology-Driven Syntax: A Theory of V to I Raising and Pro-Drop*. Benjamins, Amsterdam.
- Saltarelli, M. 1988. *Basque*. Croom Helm, London.
- Sano, T. 1995. Roots in Language Acquisition: A Comparative Study of Japanese and European Languages, Ph.D dissertation, University of California Los Angeles.
- Sanoudaki, I. 2002. Greek 'strong' pronouns and the delay of Principle B effect. *Reading Working Papers in Linguistics* 7: 103-124.
- Santelmann, L. 1995. The Acquisition of Verb Second in Child Swedish: Continuity of Universal Grammar in Wh-questions, Topicalisations and Verb Raising. Ph.D dissertation, Cornell University.
- Saussure de, F. 1949. *Cours de Linguistique Général*, fourth edition. Paris.
- Schaeffer, J. 1990. The Syntax of the Subject in Child Language: Italian Compared to Dutch. M.A. thesis, University of Utrecht.
- Schutze, C. and K. Wexler. 1996. Subject case licensing and English Root Infinitives. *Proceedings of the Boston University Conference on Language Development*. BUCLD 20: 670-682. Cascadilla Press, Somerville MA.
- Schwanenflugel, P. J. & M. Rey. 1986. The relationship between category typicality and concept familiarity: evidence from Spanish- and English-speaking monolinguals. *Memory and Cognition* 14: 150-163.
- Serratrice, L., A. Sorace and S. Paoli. 2004. Crosslinguistic influence at the syntax-pragmatics interface: subject and object in Italian-English bilingual and monolingual acquisition. *Bilingualism: Language and Cognition* 7: 1-23.
- Sigurjonsdottir, S. 1992. Binding in Icelandic: Evidence from Language Acquisition. Ph.D diss. University of California Los Angeles.
- Slobin, D. 1977. Language change in childhood and history. In J. Macnamara (ed.) *Language learning and thought*. Academic Press, New York.
- Slobin, D. 1984. *The Crosslinguistic Study of Language Acquisition (Vols. 1 and 2)*. L. E. A., Hillsdale, NJ.
- Smith, N. V. 1973. *The Acquisition of Phonology: A Case Study*. Cambridge University Press, Cambridge.
- Smith, N. V. 2003. Are gucks mentally represented? *Glott International* 7: 164-166.
- Speas, M. 1993. Null arguments in a theory of economy of projections. *Umass Occasional Papers in Linguistics*. Amherst, Ma.
- Speas, M. 2006. Economy, agreement, and the representation of null arguments. In P. Ackema, P. Brandt, M. Schoorlemmer and F. Weerman (eds.) *Arguments and Agreement*. Oxford University Press, Oxford.

- Sperber, D. and D. Wilson. 1995. *Relevance: Communication and Cognition*. Blackwell, Oxford.
- Sportiche, D. 1993. Clitic constructions. In: J. Rooryck and L. Zaring (eds) *Phrase Structure and the Lexicon*. UCLA
- Sportiche, D. 1999. Pronominal clitic dependencies. In Henk van Riemsdijk (ed.) *Clitics in the Languages of Europe*. Mouton de Gruyter, Berlin.
- Steele, S. 1975. Is it "possible"? *Stanford University Working Papers on Language Universals*. 18: 35-38.
- Stephany, U. 1986. Modality. In P. Fletcher and M. Garman (Eds.), *Language acquisition: Studies in first language development*. Cambridge University Press, Cambridge.
- Stromswold, K. 1990. Learnability and the acquisition of auxiliaries. Ph.D. dissertation, MIT Cambridge.
- Sun Song, M. 1999. The acquisition of negative polarity items by English-Speaking children. *University of Hawai'i at Manoa Working Papers in Linguistics*. 31: 75-86.
- Sweetser, E. 1990. *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure*. Cambridge University Press, Cambridge.
- Tamburelli, M. 2004. On the acquisition of lexical properties. In Yanez Bouza, N. eds. *Proceedings of the 13th Postgraduate Conference in Linguistics*, University of Manchester.
- Taraldsen, K. T. 1978. On the NIC, Vacuous Application and the That-Trace Filter. Ms., MIT.
- Tomioka, S. 2003. The semantics of Japanese null pronouns and its cross-linguistic implications. In K. Schwabe and S. Winkler (eds.), *The Interfaces: Deriving and Interpreting Omitted Structures*. Benjamins, Amsterdam.
- Torrens, V. 1992. The acquisition of inflection in Spanish and Catalan. *MIT Working Papers in Linguistics* 26: 451-472.
- Tsimpli, I.-M. and A. Roussou. 1996. Negation and polarity items in Modern Greek. *The Linguistic Review* 13: 49-81.
- Tsimpli, I.-M. 1996. *The Prefunctional Stage of Language Acquisition*. Garland, New York.
- Unsworth, S. 2003. Testing Hulk & Müller (2000) on crosslinguistic influence: Root Infinitives in a bilingual German/English child. *Bilingualism: Language and Cognition* 6: 143-158.
- Uriagereka, J. 1995. Aspects of the syntax of clitics in Western Romance. *Linguistic Inquiry* 26: 79-123.

- Vainikka, A. 1994. Case in the development of English syntax. *Language Acquisition* 3: 257-325.
- Valian, V. 1990. Logical and psychological constraints on the acquisition of syntax. In L. Frazier and J. de Villiers (eds) *Language Processing and Language Acquisition*. Kluwer, Dordrecht.
- Vallduvi, E. 1987. The discourse function of pro-drop and subject postposition in Catalan. *Penn Review of Linguistics* 11: 129-136.
- van Kampen, J. to appear. Language specific bootstraps for UG categories. *International Journal of Bilingualism*.
- Varlokosta, S. 2000. Lack of clitic pronoun distinctions in the acquisition of Principle B in child Greek. *Proceedings of the Boston University Conference on Language Development*. (BUCLD 24) 738-748. Cascadilla Press, Somerville MA.
- Varlokosta, S. 2001. On the acquisition of pronominal and reflexive reference in child Greek. In M. Almgren, A. Barreña, M.-J. Ezeizabarrena, I. Idiazabal & B. MacWhinney (eds.) *Research on Child Language Acquisition*. 1383-1400. Cascadilla Press, Somerville MA.
- Varlokosta, S. and J. Dullaart. 2001. The acquisition of pronominal reference by Greek-Dutch bilingual children: evidence for early grammar differentiation and autonomous development in bilingual first language acquisition. *Proceedings of the Boston University Conference on Language Development*. (BUCLD 25) 780-791. Cascadilla Press, Somerville MA.
- Volterra, V. and T. Taeschner. 1978. The acquisition and development of language by a bilingual child. *Journal of Child Language* 5: 311-326.
- Watanabe, A. 2004. The genesis of negative concord: syntax and morphology of negative doubling. *Linguistic Inquiry* 35: 559-612.
- Weerman, F. and J. Evers-Vermeul. 2002. Pronouns and case. *Lingua* 112:301-338.
- Wells, G. 1979. Learning and using the auxiliary verb in English. In V. Lee (ed.), *Language Development*. Open University Press, London.
- Wells, R. S. 1947. Immediate constituents. *Language* 23: 81-117.
- Weverink, M. 1990. The subject in relation to inflection in child language. M.A. thesis, University of Utrecht.
- Wexler, K. 1990. Optional infinitives, head movement and the economy of derivations in child grammar. Paper presented at the *Annual Meeting of the Society of Cognitive Science*, MIT.
- Wexler, K. 1994. Optional Infinitives, Head Movement and the Economy of Derivation. In D. Lightfoot and N. Hornstein (eds.), *Verb Movement*. Cambridge University Press, Cambridge.



- Wexler, K. 1996. The development of inflection in a biologically based theory of language acquisition. In M. L. Rice (ed.), *Toward a Genetic of Language*, Lawrence Erlbaum Associates, Mahwah, New Jersey.
- Wexler, K. 1998. Very early parameter setting and the unique checking constraint: A new explanation of the optional infinitive stage. *Lingua* 106: 23-79.
- Wexler, K. 1999. Maturation and growth of grammar. In W. C. Ritchie and T. K. Bhatia (eds.) *Handbook of Language Acquisition*. MIT Press, Cambridge, Ma.
- Wexler, K. and R. Manzini. 1987. Parameters and learnability in Binding Theory. In Roeper T. and E. Williams (eds.) *Parameter Setting*. Reidel, Dordrecht.
- Wexler, K. and Y. Chien. 1985. The development of lexical anaphors and pronouns. *Papers and reports on child language development*, 24. Stanford University, Stanford, Calif.
- Wijnen, F., M. Kempen, and S. Gillis. 2001. Root infinitives in Dutch early child language. *Journal of Child Language*, 28, 629-660.
- Williams, E. 1981. Language acquisition, markedness, and phrase structure. In S. L. Tavakolian (ed.) *Language acquisition and linguistic theory*. MIT Press, Cambridge, Mass.
- Wong, A. M.-Y., L. B. Leonard, P. Fletcher, S. F. Stokes. 2004. Questions without movement: a study of Cantonese-speaking children with and without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research* 47: 1440-1453.
- Yang, C. 2002. *Knowledge and learning in natural language*. Oxford University Press, New York.
- Yang, C. 2004. Universal grammar, statistics, or both. *Trends in Cognitive Sciences*, 8, 451-456.
- Yip, V. and S. Matthews. 2000. Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition* 3: 193-208.
- Zanuttini, R. 1991. Syntactic Properties and Sentential Negation: A Comparative Study of Romance Languages. Ph.D dissertation, university of Pennsylvania.
- Zanuttini, R. 1995. *Negation and Clausal Structure: A Comparative Study of Romance Languages*. Oxford University Press, Oxford.
- Zanuttini, R. 1997. Negation and verb movement, in L. Haegeman (Ed.) *The New Comparative Syntax*. Longman, 214-245. London and New York
- Zwanziger, E., S. E. M. Allen and Fred Genesee. 2005. Crosslinguistic influence in bilingual acquisition: subject omission in learners of Inuktitut and English. *Journal of Child Language* 32: 893-909.

- Zwicky, A. 1977. *On Clitics*. Indiana University Club, Bloomington.
- Zwicky, A. M. 1985. Clitics and Particles. *Language* 61: 283-305.
- Zwicky, A. M. and K. Pullum. 1983. Cliticisation vs. Inflection: English N'T. *Language* 59: 502-513.

## Appendix

### 1. Sample datasheet

Data: \_\_\_\_\_

Nome: \_\_\_\_\_

DoB: \_\_\_\_\_

Età: \_\_\_\_\_

1	Questo è il mago Silvan. Questo è l'aereo magico. Il mago Silvan è seduto sull'aereo?	Y/N	CONM 8
2	Questo è Superman. Questa è la supermoto. Superman sta lavando la supermoto?	Y/N	CONX 4
3	Questa è la mamma. Questa è la bambina. La bambina la sta asciugando?	Y/N	NCX 2A
4	Questo è il vigile. Questo è Pierino. Pierino lo sta lavando?	Y/N	NCM 1D
5	Questo è il pistolero. Questo è Furia il cavallo. Il pistolero sta spingendo Furia il cavallo?	Y/N	CONX 9
6	Questa è la mamma. Questa è la bambina. La mamma sta asciugando lei?	Y/N	NPM 2D
7	Questo è lo sceriffo. Questo è il pompiere. Il pompiere lo sta pungendo?	Y/N	NCX 3B
8	Questo è il vigile. Questo è Pierino. Pierino sta lavando lui?	Y/N	NPX 1A
9	Questo è il postino. Questa è una moto. Il postino sta lavando la moto?	Y/N	CONX 9
10	Questo è il pistolero. Questo è Furia il cavallo. Furia il cavallo sta morsicando il pistolero?	Y/N	CONX 9
11	Questa è la mamma. Questa è la bambina. La mamma la sta asciugando?	Y/N	NCM 2D
12	Questo è lo sceriffo. Questo è il pompiere. Lo sceriffo lo sta pungendo?	Y/N	NCX 3A
13	Questo è il vigile. Questo è Pierino. Il vigile sta lavando lui?	Y/N	NPM 1C
14	Questa è la mamma. Questa è la bambina. La mamma sta asciugando lei?	Y/N	NPX 2B
15	Questo è Shumacherino. Questo è un rimorchio. Shumacherino sta pulendo il rimorchio?	Y/N	CONX 5
16	Questo è lo sceriffo. Questo è il pompiere. Il pompiere lo sta pungendo?	Y/N	NCM 3C
17	Questo è il vigile. Questo è Pierino. Pierino sta lavando lui?	Y/N	NPM 1D

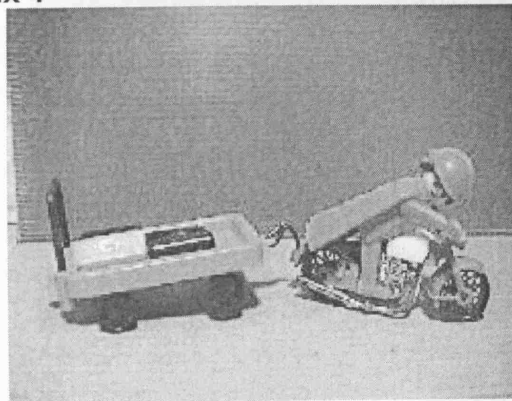
18	Questa é la mamma. Questa é la bambina. La bambina sta asciugando lei?	Y/N	NPX 2A
19	Questo é Superman. Questo é un rullo. Superman sta pulendo il rullo?	Y/N	CONX 6
20	Questo é lo sceriffo. Questo é il pompiere. Il pompiere sta pungendo lui?	Y/N	NPX 3B
21	Questo é il mago Silvan. Questo é l'aereo magico. Il mago Silvan sta spingendo l'aereo magico?	Y/N	CONX 8
22	Questo é il vigile. Questo é Pierino. Il vigile lo sta lavando?	Y/N	NCM 1C
23	Questa é la mamma. Questa é la bambina. La mamma la sta asciugando?	Y/N	NCX 2B
24	Questo é il postino. Questa é una moto. Il postino sta spingendo la moto?	Y/N	CONX 7
25	Questo é il vigile. Questo é Pierino. Il vigile lo sta lavando?	Y/N	NCX 1B
26	Questo é lo sceriffo. Questo é il pompiere. Il pompiere sta pungendo lui?	Y/N	NPM 3C
27	Questo é Shumacherino. Questo é un rimorchio. Schumacherino sta pitturando il rimorchio?	Y/N	CONX 5
28	Questo é Superman. Questa é la supermoto. Superman sta pulendo la supermoto?	Y/N	CONX 4
29	Questa é la mamma. Questa é la bambina. La bambina la sta asciugando?	Y/N	NCM 2C
30	Questo é il vigile. Questo é Pierino. Pierino lo sta lavando?	Y/N	NCX 1A
31	Questo é lo sceriffo. Questo é il pompiere. Lo sceriffo sta pungendo lui?	Y/N	NPM 3D
32	Questo é Superman. Questo é un rullo. Superman sta pitturando il rullo?	Y/N	CONX 6
33	Questo é lo sceriffo. Questo é il pompiere. Lo sceriffo lo sta pungendo?	Y/N	NCM 3D
34	Questo é il vigile. Questo é Pierino. Il vigile sta lavando lui?	Y/N	NPX 1B
35	Questa é la mamma. Questa é la bambina. La bambina sta asciugando lei?	Y/N	NPM 2C
36	Questo é lo sceriffo. Questo é il pompiere. Lo sceriffo sta pungendo lui?	Y/N	NPX 3A

## 2. Experimental Stimuli:

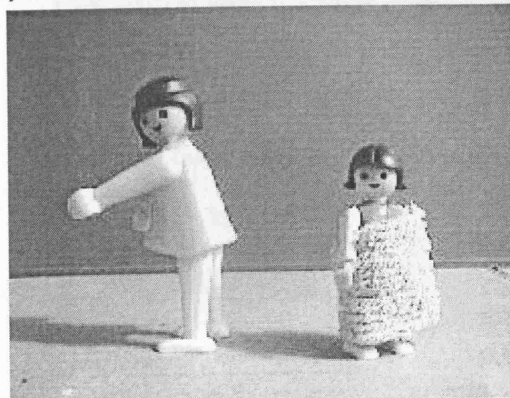
ConM 8



ConX 4



NCX/NPX 2A



NCM/NPM 1D



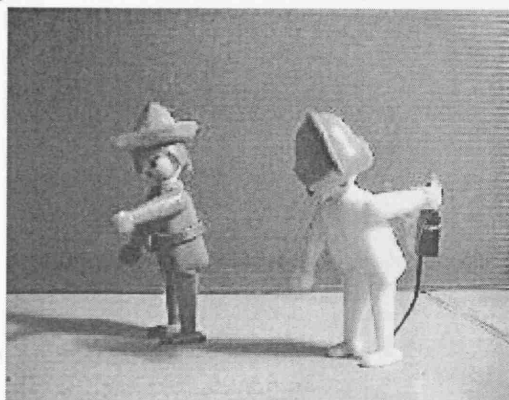
ConX 9



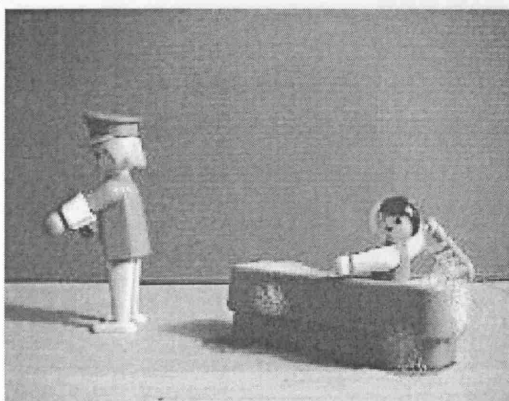
NCM/NPM 2D



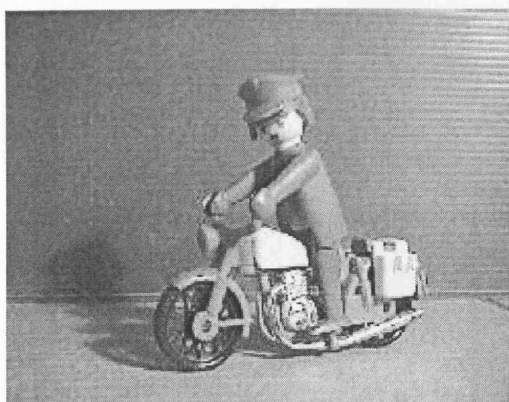
**NCX/NPX 3B**



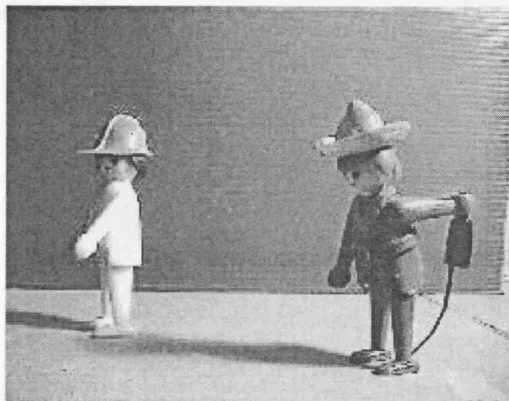
**NCX/NPX 1A**



**CoNX 7**



**NCX/NPX 3A**



**NCM/NPM 1C**

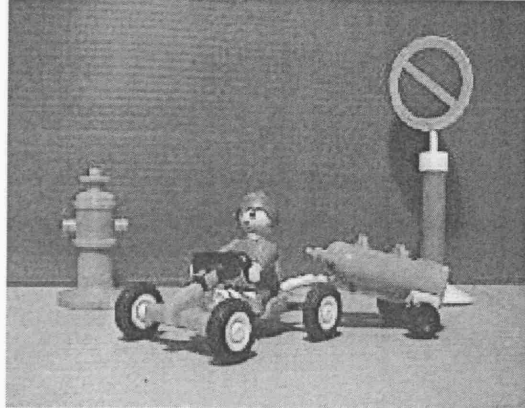


**NCX/NPX 2B**

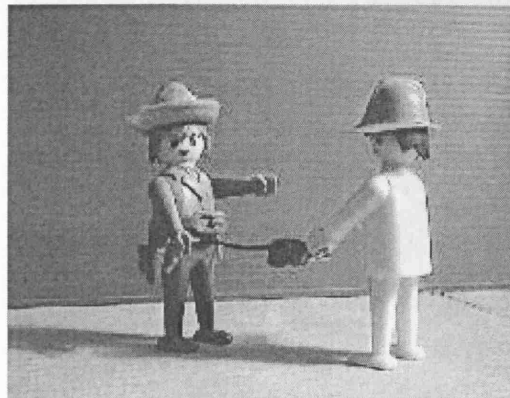




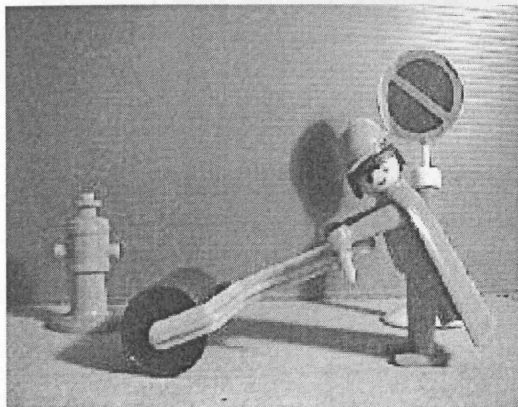
ConX 5



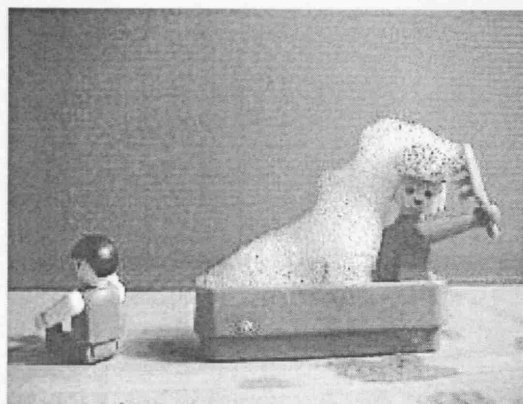
NCM/NPM 3C



ConX 6



**NCX/NPX 1B**



**NCM/NPM 2C**



**NCM/NPM 3D**

