POST-CRITICAL PERIOD AGE OF ARRIVAL AND ITS RELATIONSHIP TO ULTIMATE ATTAINMENT IN A SECOND LANGUAGE

Catherine Mary Rice

PhD

University of Edinburgh
1999

to my family

ACKNOWLEDGEMENTS

DECLARATION

I declare that this thesis is my own work.

CONTENTS

ABSTRACT

The effect of an individual's age of arrival in the second language community on her ultimate attainment in the target language has come under considerable scrutiny by Second Language Acquisition researchers in the last 25 years. However most research has focussed on comparisons of learners who arrive as children and those who arrive as adults, post-puberty. There is now a compelling body of empirical evidence that the former are much more likely than are adults to reach native levels both of performance (in phonology and syntax) and competence - that is to achieve a mental representation of the second language grammar that matches that of a native speaker. Nevertheless, many cases of successful language learning by adults are attested in the literature. Moreover not only exceptionally talented but also highly proficient adult learners reveal knowledge of the second language that cannot have been arrived at by deductive processes alone: access to the principles and parameters of Universal Grammar is clearly implicated.

The purpose of this thesis is to attempt a synthesis of age studies with ultimate attainment studies by investigating the effect of age of arrival on learners' knowledge of English as an L2 *after* puberty. Studies by Birdsong (1992) and White and Genesee (1996) conclude that there is little or no difference between adult and child intuitions: therefore there is no inevitable barrier to late language acquisition as proposed by the critical period hypothesis. Birdsong (1992) however also suggests that older adults are less likely to achieve native-like competence than younger adults. This study aims to test these conclusions.

The results of an experimental study comparing the grammatical intuitions of highly proficient Italian learners of English, both child and adult 'arrivers', and native speakers of English are presented. The results of a Grammaticality Judgement test of knowledge of wh-movement constructions show that adult and child learners can show equally native-like intuitions for invariant Subjacency and the ECP violations. Significant differences between adult and child learners were found only where there is parameterized variation between English and Italian. No differences were found between older and younger adults.

It is concluded therefore that adults do have access to UG when acquiring a second language, and that maturational factors only play a secondary role. Age after the critical period is not a factor in acquisition, at least before middle age. Thus the adult arrivers were as capable as the

child arrivers of correctly rejecting ungrammatical wh-movement sentences, where English and Italian instantiate the same constraints. Full acquisition of the target grammar, including those properties where the L1 and L2 settings vary parametrically appears to be blocked by prior knowledge of the first language.

*

CHAPTER 1

INTRODUCTION

1.1   Variability

1.2   The Age Factor

1.3.   Universal Grammar and Second Language Acquisition

1.4.   The Present Study

1.5   Summary of the Thesis

1.6   Terminological Note and Glossary

1.1   Variability

The most glaring, most interesting and perhaps the most puzzling difference between native speakers of a language and those who acquire it as a second language is the difference in ultimate attainment. Nearly all humans - even those suffering from many forms of deprivation and disability - learn to speak and understand their first language fluently and accurately, although there may be considerable differences in lexical breadth and grammatical complexity in performance, not to mention dialectal and idiolectal variation, both between native adult speakers of the same language and within the spoken utterances of an individual. Moreover, most adult speakers share similar intuitions about what can and cannot be said (what is grammatical and what is ungrammatical) in their native language. The theory of Universal Grammar (UG) holds that a "number of fundamental principles that sharply restrict the class of attainable grammars and narrowly constrain their form, but with parameters that have to be fixed by experience" (Chomsky 1981:3-4) constitutes human knowledge of language and enables humans (uniquely) to acquire language as described above. Children thus acquire language not by induction or by analogy, but through the indirect action of the utterances they hear on the principles and parameters of UG, already configured in the language faculty of the brain.

At this point, it is necessary to introduce the distinction formulated by Chomsky (1965) between *competence* and *performance* in a language. Competence refers to latent or tacit knowledge of a language, which allows adult speakers of the language to decide what is and what is not grammatical (Newmeyer 1983). Performance refers to the spoken utterances, or written productions, of individual speakers, which may be more or less grammatical, complete or comprehensible at different times depending on a wide range of psycholinguistic, pragmatic, and physiological factors. Thus two native speakers of the same language (L1) may be totally dissimilar in their speech productions, but are likely to agree in their intuitions about the grammar of their language. Equally, the speech performance of an individual may vary in grammatical accuracy, as well as in lexical choice or stylistically, from one occasion to another, yet her knowledge remain unchanged.

Most second language (L2) speakers who learned their additional language as adults can be identified as such simply by their lack of success in reaching a native-like level of

proficiency in accent, morphology, syntax, lexis or discourse. Child learners on the other hand are much more successful. Only a very small minority, perhaps 5% (Selinker 1982), achieve native-like levels of fluency and accuracy in a wide range of situations. Whether any of this elite ever match native-like competence in their second language is still an unresolved question, and the subject of a small but significant body or research (Coppieters 1987; Birdsong 1992; Sorace 1992; Ioup, Boustagui, El Tigi and Moselle 1994, White and Genesee 1996).

The reasons for these differences between L1 and L2 speakers, and amongst L2 speakers, are a central concern of research into second language acquisition (SLA). Non-linguistic factors such as aptitude, psycho-social factors and methods of instruction have received considerable attention over the last 30 years, as have linguistic and psycholinguistic variables such as distance between L1 and L2 and the availability of Universal Grammar in second language acquisition. Age as a factor in the variability in ultimate attainment of adult L2 speakers has however received rather less attention. It is the purpose of this thesis to present data from an experimental investigation into the influence of age at first exposure to a second language, in particular after adolescence, and to assess it as a cause of differences in ultimate attainment.

1.2 The Age Factor

Neither motivation, social "distance", aptitude nor method of instruction appears to play any part in first language acquisition (FLA). Children acquire their native language regardless of such environmental or voluntary conditions. All children moreover (including deaf and blind children and children suffering from Down's syndrome and other disabilities) who are exposed to language will learn it "completely" at much the same age, with minimal intervention by adults. It would appear that only if a child is deprived of normal exposure until adolescence is it likely that she will fail to acquire language (Long 1990). Age can, on the other hand, help to reduce language proficiency in later life: old people often speak and understand a more simplified linguistic code than younger adults, showing, in a number of studies, a reduced capacity for using and processing complex sentences (Feier and Gerstmann 1980, Emery 1985, Kemper 1986, Kynette and Kemper 1986, Obler 1993,

Bloom, Mullins and Paternostro 1996).

In second language acquisition (SLA) research, the influence of the age at which an individual is first exposed to a second language has been investigated from both socio-psychological and maturational standpoints. Most researchers however have been concerned to discover whether children who are exposed to their second language before adolescence (during the critical or sensitive period, first described by Lenneberg (1967)) are privileged with regard to their ultimate attainment as compared with adults. There has been very little in the way of comparisons of the ultimate attainment in performance or competence of adult age groups.

1.2.1 Psycho-social explanations

Sociolinguistic and psychosocial studies have shown that older learners on the whole are less likely to be motivated to integrate into the L2 community than younger ones, and that this will have a marked effect on the amount of L2 input they receive, and hence on their rate of acquisition and their final level of proficiency (Kessler and Idar 1979, Schumann 1978, Cancino 1985). Children probably get more exposure to the L2, may have more compelling reasons for integration and may find "acculturation" - becoming part of the target culture (Schumann 1978) - easier. Grosjean (1982:179) states:

> [T]he degree of bilingualism attained is not related to whether the languages are acquired
> simultaneously or successively. It is psychosocial factors, such as the use of the language
> in the family or in the school, that will condition when, to what extent, and for how long
> a child will be bilingual, not the age of acquisition of the two languages.

After adolescence and the end of formal schooling, learners, particularly those who have migrated for economic or political reasons to another country, are more likely to prefer to maintain links with their L1 community, or on the other hand to find it harder to assimilate with the L2 community. The cluster of behaviours associated with such "social distance" has been held to account for fossilization, whether of syntax and lexis at an early stage, or of accent among highly proficient second language speakers (Cazden, Cancino, Rosansky and Schumann 1975, Schumann 1978).

In old age, many bilinguals have been observed to prefer to reintegrate with their L1 community if they can, and reduce their contact with the L2 (see De Bot and Clyne 1989).

It seems clear from the above that sociocultural and psychosociological variables can explain much of the variability in ultimate attainment between child and adult learners. Yet there remain both individual cases and areas of variability which are not amenable to psychosocial accounts. Grosjean (1982:179) argues that there is little difference in final achievement between children brought up bilingually from birth by bilingual parents (simultaneous acquisition), and those (usually from migrant families) who acquired their second language at school (successive acquisition), sometimes in harsh and alien circumstances. What matters, he maintains, are psychosocial factors, related to quantity and quality of input. Two sets of evidence run counter to this position: motivation has not been found to play a significant part in ultimate attainment amongst children (Johnson and Newport 1989, Oyama 1978); amongst adults there are cases where highly motivated and integrative learners have fossilized at an early stage (Schmidt 1981). Moreover, as Long (1990) points out, the claims made for the impact of these variables have not been precise, or tested empirically: "Just which of these variables, in what combinations, and to what degree, are supposed to affect learning and why?" (Long 1990:275).

1.2.2. Maturational aspects

Comparisons of neurological or maturational aspects of the rate of acquisition, ultimate achievement and grammatical knowledge of child and adult learners have clearly established that the former are more likely to reach native-like proficiency than adults, although not necessarily so quickly, or so efficiently. There are however many instances of highly proficient and near-native adult learners who would appear to challenge this prevailing tendency. This phenomenon raises two key questions: (1) What are the neurological features, skills or circumstances that allow such achievement? (2) Does native-like proficiency in adults necessarily entail native-like competence? The first question relates to the conditions for successful learning and does not directly concern us here. It has been the subject of important investigations which will be touched on later. The second question raises further issues: is there a difference between the mental representations of the L1 grammar held by native speakers and by native-like L2 speakers, or are they the same? If different, is there a bilingual competence which is different from a monolingual competence (Cook 1991:115)? Findings from a number of studies give diametrically opposite answers to these questions.

Some writers have proposed that only the immature brain can, when exposed to a second language, eventually produce a native-like competence (Johnson and Newport 1989,1991). Others have pointed to L2 speakers who, despite acquiring the L2 in adulthood, do not differ from native speakers in their intuitions about the target language (Birdsong 1992, Ioup et al 1994, White and Genesee 1996). Still others maintain that although adults do not attain the same native-like competence as children, their mental representations of the target grammar are subject to universal constraints and their L2 knowledge is more than could be deduced from the language input they receive (White 1992, Sorace 1993). A final question is raised by this last position: might adult speakers who are highly proficient but not native-like in their oral production also share some intuitions with native speakers?

1.3. Universal Grammar and Second Language Acquisition

If there is a maturational effect in second language acquisition, the question of the availability of Universal Grammar seems likely to be intimately connected to it. While Chomsky's formulations have not generally attempted to deal with the question of the *process* of child language acquisition, a theory postulating an innate "language instinct" may be the most parsimonious explanation for the data on maturational constraints on language learning, although clearly not the only one (eg Piaget's cognitive development model). Universal Grammar[1] refers to the theory that there is a set of innate principles and parameters with which everyone is endowed, and which enables children to acquire their first language in the teeth of insufficient and often contradictory input from the utterances of those around them, and in the absence of any consistent negative input (instruction, correction, metalinguistic information) from adults (Birdsong 1989). Universal Grammar describes the "finite system of knowledge which enables [human beings] to construct and interpret an infinite number of sentences" (Haegeman 1994:9). As far as acquisition is concerned, UG proposes that the role of the language environment is to trigger a sequence of innate constraints which children can be shown to observe from their earliest speech development. If this sequence is maturationally directed, then there should be an intimate relationship between maturation and the acquisition of the principles and parameters of UG.

The question that researchers into SLA have posed is: is UG also available to the second language learner? Second languages are generally - except for children brought up by parents with two different first languages - acquired in very different conditions from those of first languages. Adults have knowledge of the L1 and the experience of learning it and they may also receive plentiful negative evidence in the form of formal instruction, although as many studies have shown, negative evidence may be ignored even here, and is also largely unavailable for naturalistic learners. Uneven attainment aside, many learners do achieve a proficiency in the L2, not necessarily native-like, which is not easily explained by inductive processes alone (Birdsong 1989, White 1992, Schwartz and Gubala-Ryzak 1992).

As we have already noted, there are no cases of fossilization in first language acquisition (FLA). The "steady state" achieved by adults is a complete mental representation of the grammar. This is attested across a wide range of IQs and amongst children brought up in "abnormal" circumstances, such as by deaf parents or in isolation (Long 1990). The UG theory of acquisition has to cope with the fact that only positive evidence is reliably available for the core grammar. Correction and instruction by parents (negative evidence), though much of the evidence is anecdotal, appears to be inconsistent, unfocussed and largely ignored (see Birdsong 1992:ch.5). Even when positive evidence is seriously impoverished and incomplete, children still develop speech normally.

According to the Maturation Hypothesis (Borer and Wexler 1987) UG matures like other physical and neurological organs. This could explain why triggering data, present in the evidence from the start, takes time to activate parameter settings: why, for example, English-speaking children persist with null subjects for so long, in the face of positive evidence of expletives, or verb inflections. The major question of FLA - how to explain children's success when negative evidence is largely absent - is tackled by the Subset Principle, which states that children adopt the least marked (most conservative, that is, the subset) hypothesis consistent with the evidence:

> If acquisition can proceed using only positive data, then it would seem completely unnecessary to move to an enrichment of the input data that is as yet unsupported by psycholinguistic evidence. Beyond this point, limiting the procedure to positive evidence does not hinder learning - it helps, since we can focus on the important matter of constraints needed so that learning can proceed. (Berwick 1985:85)

The Subset Principle depends on the Subset Condition (Wexler and Manzini 1987), which states that the grammars produced by different settings of a parameter are in a subset-superset relation to each other. The child selects the smallest language compatible with the input, and will only reconstruct her hypothesis when confronted by disconfirming positive evidence.

Intermediate grammars (ie those that are different from adult native grammars) thus do not, according to this and other learning theories, violate UG constraints: what happens is that children adopt "foreign" but not ungrammatical settings, as in the case of null subjects (Hyams 1987).

Although the Principles and Parameters framework developed by Government-Binding (GB) theory (Chomsky 1981) was not formulated to take account of second language acquisition, it contains several features which make it attractive to SLA research. It starts from the position that there is a universal human grammar, and that variation between different languages is very limited. By proposing a small number of universal constraining principles rather than sets of rules (as in earlier versions of generative grammar), the theory seems to invite application to second language learning: if first language learning is simply a question of activating principles and setting parameters in response to input, this should also be the most economical explanation in the case of learning a second language. There are other similarities. Plato's problem holds in SLA as well: "L2 learners often end up with a highly complex unconscious mental representation of their second language (not necessarily identical to a native speaker's grammar), which is underdetermined by the L2 input, suggesting that built-in knowledge must be involved" (White 1992:8-9). The role of negative evidence seems almost as questionable as in L1A. Such similarities as well as the inherent parsimony of a UG-based model of L2A have stimulated, through the 1980s and 1990s, a flood of research.

Most of this research has concentrated on the acquisition process rather than on the final product (although there is still no complete model even of child L1 acquisition) and has been concerned to determine to what extent UG is available in SLA. Four positions can be identified (White 1989,1992; Young-Scholten 1994):

(1) UG is only available in FLA and child SLA, and second languages are acquired by adults through general cognitive processes.

(2) UG is indirectly accessible via the L1; parameters set for the L1 cannot be reset for the L2.

(3) Adult and child L2 learners have the same competence as native speakers and this is achieved through the operation of UG principles, as in FLA. In setting parameters for the target language, the learner "returns" to UG, and follows much the same route as a child acquiring the same language.

(4) The competence of native and adult non-native speakers is different, though both are constrained by UG. Due to learnability issues the Interlanguage non-native grammar may not resemble either the L1 grammar or the L2 grammar.

A problem that besets this field is that experiments are rarely comparable: procedural shortcomings in selection of subjects, choice of grammatical features, test administration and in the analysis of results make it often difficult to replicate experiments or to make comparisons across studies. However Zobl (1992) compared six recent studies which tested subjects' intuitions on the grammaticality or otherwise of sentences. Subjects were advanced learners of different second languages. He found that intuitions were more homogeneous on UG-violations than on sentences requiring data-dependent (ie language-specific) knowledge. These results imply that processing constraint violations is less complex and less demanding because well-formed counterparts are not easy to compute. Hence, Zobl argues (1992:398) "where knowledge derives from innate mental structure without the interaction of experience, performance should be more regular across the species".

In support of the first position summarised above, Schachter (1988) argues that there are too many dissimilarities between first and second language acquisition to make the availability of UG in adult SLA a reasonable assumption. She points to 4 major areas of difference. First, very few adults achieve complete, native-like knowledge of a second language. What they end up with is a very wide range of proficiencies "ranging from barely communicative to highly fluent, with most second language speakers fairly evenly distributed along the continuum that might be labelled as one of communicative competence" (224). Secondly, unlike children, not all adults have an equal potential for learning languages. The degree of relatedness between her first and second languages also has a large influence on an individual's success in acquiring the second language. UG-oriented researchers acknowledge this: Zobl (1990) points out that his Japanese students made few mistakes with referential

pronoun subjects in English compared with the Spanish learners investigated by Phinney (1987). This difference he accounts for by the fact that Spanish learners of English need to reset the parameter but Japanese learners need to activate it, as Japanese does not instantiate the agreement parameter. However, the relationship between a learner's L1 and L2 is more complex and less linear than this single example suggests: Sorace (1993) shows that L2 learners of Italian whose first language instantiates a principle of the L2 (unaccusativity) are more accurate in the L2 than learners whose L1 lacks this principle. In the area of phonology there is evidence that the closeness of L1 to L2 does not necessarily promote easier or more complete learning (Flege 1987, Wode 1994). Finally, fossilization, which never happens in FLA, either as the overt persistence into adulthood of early forms, or the re-appearance in emotional and heightened situations of superseded forms, is a common feature of interlanguages. Although Schachter's article does not clearly differentiate between core and periphery grammatical knowledge and so seems often to be arguing against a non-existent position that would attribute the whole of SLA to the operation of UG, it does raise issues which, as we shall see later, have yet to be satisfactorily resolved.

Properties of language which now appear to be language-specific may with the further development of the theory prove to be part of UG. An example from studies of production data by "naturalistic" acquirers of an L2 is relevant here. Clahsen (1988) and Clahsen and Muysken (1988) argued, from a study of the acquisition of negation by naturalistic learners of German, that the learners do not observe UG constraints, and produce "unnatural" grammars. Tomaselli and Schwartz (1990) reanalysed the data on the basis of recent work on verb movement by Pollock and others and concluded that the learners do have access to UG, and apply UG principles throughout the three stages of negation elaborated by Clahsen and Muysken. However, UG is not an empirically verifiable fact, only a theory, and as Tomaselli and Schwartz point out, "it is always possible to come up with a non-UG-based analysis of L2 data"(1990:27). UG-based analysis is not necessary, but it is the most "feasible" explanation (Lust 1988:311). Lust also suggests that although a data-driven, inductive route to L2 acquisition would seem more apt for adult learners, precisely because of their previous experience of language learning, the evidence is that they seem not to take such a route.

While there is a large body of evidence to support the theory that innate language knowledge is at work in child acquisition of a first and even a second language, in adult SLA the claim that UG is available is much more contentious. However, there are many studies which tend

that UG is available is much more contentious. However, there are many studies which tend to show that UG plays a role of some kind in adult acquisition. A much more interesting question now is: what kind of role? Again, to look at the question from another angle, can a UG-based account of SLA explain incomplete and uneven attainment? The fourth option (see above: (different competence, same mechanism) with UG interacting with the LI grammar and with the input according to learnability considerations allows for a wider variety of UG-constrained acquisitional outcomes. Whether age as a variable interacts with UG to afford greater or lesser access to it at different points in the life-span is the main locus of interest of this investigation.

1.4.   The Present Study

So far, research into this area of second language acquisition has focussed on two-way "before and after" comparisons of the metalinguistic judgements of highly proficient or near-native child and adult learners, with puberty (variously taken as 14,15,16 or 18) as the watershed. There has been little in the way of comparative studies of the ultimate attainment of older and younger adults (Birdsong 1992, Birdsong and Molis 1997 are exceptions). Although informants with widely differing ages of first exposure have taken part in different studies, the published results have not usually given results for different adult age-groups.

Do the chances of achieving native-like knowledge of a second language continue to decline throughout the life-span, after adolescence? Alternatively, do the neurological changes of puberty bring about a once-for-all change in language learning ability which age increments thereafter do not affect?

We present here an investigation of the grammatical intuitions of highly proficient Italian native speakers about one property of UG which is instantiated in both English and Italian: the Subjacency Condition, a constraint on question and relative clause formation. The study compares the intuitions of informants who were first exposed to English in childhood and at different ages in adulthood. We predicted that age of arrival beyond puberty would continue to affect grammaticality judgements, such that the later the age of arrival the more judgments

would differ from those of native speakers, as Birdsong concludes (1992).

In the event, experimental results showed a significant negative correlation between age of arrival and test performance between adult and child arrivers on only one, parameterised, aspect of the Subjacency Condition. It is argued that there is therefore a maturational effect for second language learning, as for first language acquisition, but that it is much less catastrophic than most advocates have proposed: no evidence was found to suggest that adult learners are unable to tap into Universal Grammar, only that access may be more restricted for adults compared to children. If the human brain's capacity for language learning does deteriorate even after the end of the critical period (hypothesized by Lenneberg (1967) to last, for language learning, from 2 to 14 years), we suggest that non-maturational neurological factors may be responsible.

This thesis attempts to show that age-effects for second language acquisition throughout adulthood deserve more investigation, and that research can both add to our empirical knowledge of the limits on and opportunities for successful second language learning and contribute towards the construction of theory in this field.

1.5   Summary of the Thesis

Chapter 2 deals with the Critical Period Hypothesis of Lenneberg (1967) and subsequent research into the extent to which first and second language acquisition is constrained by the age of first exposure. It then gives a brief sketch of research into language performance and competence in older L1 and L2 speakers with the purpose of identifying features which might influence the capacity for language learning in later life. Finally, short-term or working memory is assessed as a possible candidate for variable language learning success in adulthood.

Chapter 3 discusses the question of ultimate attainment in second language learning, and recounts five studies of ultimate attainment. It provides a rationale for testing the competence of highly proficient L2 speakers.

In *Chapter 4*, constraints on wh-movement in both English and Italian are described, and the literature of wh-movement acquisition in first and second language acquisition is summarised. Non-syntactic features of metalinguistic performance such as parsing and lexical features are also discussed.

Chapter 5 lists the research hypotheses, together with brief summaries of their theoretical motivation.

In *Chapter 6* we look at the methodologies and procedures adopted by different studies which have made use of grammaticality judgement tests, and specifically those that have tested for Subjacency. Procedures used hitherto are critically appraised and the methodology adopted for this study is presented.

Chapter 7 gives an account of the experiment undertaken for this thesis, together with the results obtained.

Chapter 8 contains a discussion of the results, how they relate to the research hypotheses and what they imply for the field of age research in second language acquisition.

Chapter 9 concludes the thesis by returning to the questions raised in the earlier chapters. These are reassessed in the light of the present experiment, and broader issues prompted by the investigation are discussed.

1.6 Terminological Note and Glossary

Throughout this thesis, the terms "adult arriver" and "adult learner" are used, in contradistinction to "child arriver" or "child learner" to denote individuals who are first exposed to naturalistic L2 input after puberty. Questions relating to the role of formal instruction in an L1 environment, often a feature of the childhood of "adult learners", are dealt with in Chapter 6.

The following acronyms are also used:

FLA = First Language Acquisition

SLA = Second Language Acquisition

UG = Universal Grammar

GB= Government and Binding

*

---

[1] The version of Universal Grammar referred to throughout this thesis is that derived from the "Government and Binding" and "Barriers" frameworks (see Chomsky 1981,1986).

CHAPTER 2

AGE AND LEARNING A SECOND LANGUAGE

2.1   The Critical Period Hypothesis

2.2   The Critical Period Hypothesis in Language Acquisition and its Critics

2.3   Second Language Acquisition

2.4   Language Abilities in Later Life

2.5   Memory and Language Learning

2.1 The Critical Period Hypothesis

The notion of a critical period for the acquisition of species-specific characteristics or behaviour is at least 100 years old. Its importance is that it "represents one of the only verifiable qualitative transitions in development" (Colombo, 1982:260). In biology and ethology, where investigations began, critical periods have been identified for the acquisition of birdsong, the socialization of dogs and binocularity in cats. What are the criteria for a critical period? Bornstein (1989) identifies 14 criterial dimensions in 4 categories - temporal contours, mechanism, consequences and time-scales. Thus a critical period has clear onset and offset times, a maturational event which takes place between these times, an identifiable external stimulus to which the organism is sensitive and - finally but crucially a "mechanism" or - in Colombo's term - a "system" which is affected by the external stimulus during the period. These feature categories are not specifications however. The temporal contours for example can be delayed, or anticipated, from individual to individual (due to environmental factors); the extrinsic stimulus may be quite weak or degraded but still trigger the maturational change. For some behaviours, indeed, it may be more accurate to talk about "optimal" or "sensitive" rather than "critical" periods: if the organism misses out on the necessary stimulation and the system fails to develop at the normal time, the behaviour may still emerge later, thanks to a reorganisation of the central nervous system or to some continuing plasticity of the system, post-critical period. "Nearly every demonstration of a critical period in behavioural development during the past 50 years has been followed by a demonstration of some behavioural recovery from the effects of critical period exposure or deprivation" (Colombo 1982:268). In what follows we use the more familiar and widely-used term "critical period", referring to both the stronger and weaker versions of the hypothesis, although several authors have preferred "sensitive period" for the latter.

As will become apparent below, finding the system that is affected by the maturational change and therefore responsible for the emergence of the relevant behaviour can be the hardest task of all, but also vital to the success of the hypothesis. In the field of language acquisition, the debate over the Critical Period Hypothesis hinges to a great extent on identifying the system: is there a component or cluster of components in the maturing brain that enables language to be acquired optimally or exclusively during a critical period?

2.2 The Critical Period Hypothesis in Language Acquisition and its Critics

Scientific investigations into the age - or more properly - the maturity factor in language acquisition began with the work of Penfield and Roberts (1959), but attracted the interest of language acquisition researchers with the publication of Lenneberg's "Biological Foundations of Language" in 1967. Lenneberg (1967) proposed that a first language can only be acquired completely if exposure to language takes place before puberty:

> Thus we may speak of a critical period for language acquisition. At the beginning it is limited by lack of maturation. Its termination seems to be related to a loss of adaptability and inability for reorganisation in the brain (179).

Children universally start to acquire language around the time they learn to walk (approximately 12 to 18 months), but do so independently of their ability to acquire other skills, cognitive or motor. By adolescence, a stable "adult" grammar and phonology are in place (with lexical acquisition continuing throughout the lifespan, albeit more slowly) and the plasticity of the immature brain (the essential ingredient in the ease with which the child mastered its first language) is lost, due, Lenneberg hypothesised, to lateralization of the brain functions. For around 95% of right-handed individuals and a majority of left-handed people, the left hemisphere is dominant for language. Using neurological evidence notably from studies of child and adult recovery from aphasia (language loss due to brain damage) and from Down's Syndrome patients, Lenneberg argued that first language acquisition after this period necessarily involves the learning of semi-linguistic communication strategies using general cognitive processes.

Lenneberg has inspired over 30 years of research, but the neurological evidence presented to date has not revealed the maturational secret behind the critical period. Krashen (1973) re-examined Lenneberg's data and concluded that lateralization or specialization of certain mental processes by left or right cerebral hemispheres starts at birth - perhaps in utero - and is complete by 4 or 5, for most individuals. First language acquisition (FLA) and possibly early second

language acquisition (SLA) is thus achieved by the left hemisphere, which is dominant for language, while the right hemisphere is generally implicated only in later SLA (and in the rare cases of late FLA) (Curtiss 1977, Schneiderman and Desmarais 1988). However, functional localization may not be so strict as Lenneberg proposed (Vargha-Khadem, Carr, Isaacs, Brett, Adams and Mishkin 1997), which suggests that the plasticity of cerebral function may derive from some other neurological features. Further studies of childhood aphasia have shown that children do not in fact recover better than adults from certain types of aphasia, and that older children can recover better than younger age-groups (Snow and Hoefnagel-Hohle 1978).

As lateralization was the only factor clearly implicated in the loss of cerebral plasticity on maturity by Lenneberg, and especially following Krashen's re-analysis of the clinical data, much work has focused on this area. Experiments with adult epileptic patients who have undergone hemispherectomies to prevent the spread of the disease have shown that the non-dominant hemisphere (usually the right, even in left-handed individuals) can comprehend very simple sentences, phrases and nouns. In the event of massive damage to the dominant hemisphere, therefore, recovery of some language function may be feasible even in adults, via utilization of the non-dominant hemisphere (McLaughlin 1984) (see below, the case of Alex). Other aspects of cerebral maturation, such as myelination - which is accomplished over decades - may be responsible for the loss of plasticity. Michael Long (1990), in an extensive review, cautiously concludes that "positing a role for neurological factors with cognitive consequences... - specifically incremental losses of plasticity with increasing brain maturation, possibly associated with myelination – if only by default seems a more defensible position [than affective and input accounts]" (Long 1990: 280). While the lateralization process does not, as we have seen, fit the period when children are acquiring language, the much longer process of myelination does. Myelination is the wrapping of axon projections of neurons in myelin sheaths of lipids and proteins and results in the specialization of neural space - in other words the functional maturation of the brain. Thus, prior to the completion of this process (which begins with the neurons implicated in physiological functions and continues with those involved in complex mental activity such as language) the necessary plasticity may be available. However, as Long points out, "the neural plasticity position is not without problems of its own... notably its lack of much empirical content and of precisely synchronized relationships with claimed losses in language learning

abilities" (280). We may conclude that not enough is known at present to pronounce definitively on the connections between plasticity loss, lateralization and the language function. And although recent developments in techniques for scanning brain activity, such as Event-Related Brain Potentials, functional Magnetic Resonance Imaging and PET, have permitted new explorations of the human brain (Ojemann 1991, Demonet, Wise and Frackowiak 1993, Zatorre 1989), ethical considerations - fortunately - restrict more invasive investigations.

There remains the linguistic evidence. Inevitably, cases of post-critical period primary acquisition are rare, and the circumstances always aberrant. Two groups of individuals have been studied in this respect however. The cases are sparse and usually incompletely reported, but some children who have been brought up isolated from a normal linguistic environment, by disability or from being deliberately hidden away by adults ("feral" or "wolf" children) have caught up in their language development so long as they were restored to a normal environment by 5-6 years (Long 1990). Where discovery was delayed until the eleventh year, the children did not develop complete proficiency, especially in closed class items such as pro-forms and auxiliaries (Curtiss 1980).

Lenneberg tended to discount evidence from accounts of feral or wolf children as unreliable for his purposes and involving too many abnormal, non-linguistic factors. However, in 1971, a thirteen-year-old girl, Genie, who had been confined to a single room from the age of 20 months and deprived of virtually all language or human contact, was discovered. The issues raised by Lenneberg's work encouraged intense interest in her case amongst linguists and it has been exceptionally well-documented. Genie has been called in evidence both for and against the Critical Period Hypothesis. After her rescue, she achieved considerable mastery of English, but it remained defective and incomplete, notably in morphology and syntax. For example, she would omit closed class items, "fragile" features which are also omitted in pidgins (Goldin-Meadow 1982:74) and which are also typical of right hemisphere language processing. Genie was "visually and tactilely oriented, excellent at remembering faces, good at finding her way around in real space" (Curtiss 1977:221) – all right hemisphere derived skills, although she was right-handed. Lateralization studies undertaken with Genie strongly indicated right hemisphere processing of both language and non-language functions. On Gestalt perception tests she performed above

average, while on sequential order (a left hemisphere function) her performance was well below; she also found counting slow and effortful. Dichotic listening tests also showed strong right hand lateralization, with results comparable to hemispherectomy subjects. Nevertheless she could communicate with language effectively. The case of Genie shows that post-critical period primary language acquisition is possible but that it involves only the right hemisphere of the brain (Curtiss 1977:234).

However, there are features of Genie's story which suggest that it is not a "straightforward" case of post-pubertal first language acquisition: she learned to speak in rather special circumstances, not wholly naturalistically but through intensive one-to-one sessions with psychologists and linguists; moreover it was not possible to discover exactly how much language she had encountered before or during her long isolation; when discovered she was 13 and still within the critical period. Finally, while she was assessed as having normal intelligence, the mental and physical suffering she endured could well have affected her language learning ability. For example, in spite of all the evidence for right hemisphere dominance for language, Genie lacked the "automatic speech" (greeting, rejoinders, swearing) considered typical of such dominance. Curtiss suggests that this "conversational incompetence" may be the result of lack of early socialization. Genie's history, which in many ways matches that of other feral children, perhaps justifies Lenneberg's omission of this type of evidence (Lamendella 1977:172).

A more recent instance of late first language acquisition which has raised equally fascinating questions is that of Alex (Vargha-Khadem et al 1997), who achieved virtually normal speech using only the right hemisphere. Alex suffered from Sturge-Weber Syndrome affecting the left hemisphere of the brain, was hemiplegic and had virtually no speech. At the age of 8;6 a left hemisperectomy was carried out to relieve seizures. At 9;6, and one month after anticonvulsant drug therapy ceased, he suddenly began to speak for the first time. Shortly before the hemispherectomy his mental age had been assessed at between 3 and 4, receptive language at the same age and productive language at 1;9. Over the next 5 years, his language development was tracked. Alex's progress in articulation, vocabulary, grammar, length and complexity of utterance and morphology was impressive over this period (eg from below 2 years to 8;5 for grammar; mean length of utterance from 0 to 11.6 words). However, some aspects of Alex's language

performance developed much less rapidly. For example, complex utterances proved difficult to process, and his ability to repeat non-words was poor, both of which, the authors speculate, may be related to short-term memory deficits. Limited short-term memory, together with low IQ (Alex's full scale IQ at 15 years was assessed at 52, with verbal IQ at 59) are found to be typical of left hemispherectomies in childhood. Phonological imitation - a measure of working memory (see Gathercole and Baddeley 1989) - and phonological awareness, a strong predictor of reading ability, were only at the level of a 5-6 year old when Alex was nearly 15. Vargha-Khadem et al (1997) suggest that it is this limited short-term memory span which represents the "cost" of right hemispherical speech[1]. They also propose that the gap between Alex's mental age (around 7 by age 15) and his language age (18 months higher) lends support to the view that language develops "somewhat independently" of cognition. On the other hand, his long-term memory showed a steady improvement (assessed at 73 on the Wechsler Memory Scale, higher than IQ). Alex represents a serious challenge to a critical period hypothesis based on lateralization, in that he achieved clear, fluent and grammatical speech well after functional lateralization.

The acquisition of signing as a first and as a second language by deaf people at different ages, ranging from infancy to adulthood, has also been a rich source of evidence in the critical period debate. Perhaps this research has raised more questions than it has answered. Newport and Supalla (1990), in a study of congenitally deaf native and later learners of American Sign Language (ASL) as a L1, found not only a strong age effect, with non-native performance beginning to become visible in some speakers from the age of 4, but also a concomitant change in learning style: native and younger learners' errors indicated that they were learning to sign by morphological analysis (their errors were caused by sequencing morphemes rather than by producing them simultaneously) while later learners made progressively greater use of unanalysed chunks and frozen forms, often combining chunks ungrammatically. All their learners had been using sign language for 10 years, so the errors did not relate to variable practice or exposure time. Similarly, a wide-ranging study by Mayberry and Eichen (1991) investigated age effects on lexical recall and on the phonological, syntactic and semantic proficiency of native child and adolescent signers. They found a sharp decrease in accuracy (in a sentence recall task) from the "native" learners (those born to deaf, signing parents) to the "child" (beginning from 5-8 years old) and "adolescent" (9-13 years old) learners, particularly at the higher syntactic and semantic levels.

This was because of the greater time spent on processing phonological aspects: late learners seemed to focus on processing surface features of utterances, requiring such effortful encoding, organising and recognising of signs that they were unable to give enough attention to meaning. "Semantic content is thus spotty, so that comprehension is incomplete" (Mayberry and Eichen 1991: 505). Nevertheless, even the late learners achieved over 50% recall rates, compared to 80% for native signers. In view of the much more rapid age-related fall in proficiency compared to second language learners, they conclude that "childhood may be more critical to first than to second language acquisition" (Mayberry and Eichen 1991:509). Conversely, one might argue that, as with the "wolf" children, the adverse social and psychological circumstances surrounding the formative years of the two older groups - unable to communicate effectively even with members of their families, denied the right to sign at schools for the deaf committed to "oralism" - cannot but have had a negative impact on the acquisition process. It is important to note that Mayberry and Eichen's bilingual signers who learned to sign in adolescence after the loss of normal hearing actually outperformed late monolingual signers. This evidence seems to support the exercise hypothesis (Johnson and Newport 1989) which predicts that acquiring a first language within the critical period is a necessary (and sufficient?) precondition for acquiring subsequent languages. Like Mayberry and Eichen (1991), Emmorey, Bellugi, Friederici and Horn (1995) conclude that "late exposure to language affects only certain aspects of linguistic processing: specifically, early phonological processing"(21). They tested adult signers (native, early and late) with on-line processing and grammaticality judgement tests, and found that only the native signers (exposed to ASL before the age of 2) were sensitive to verb agreement morphology, but that the three groups performed similarly on an offline grammaticality judgement test (and on verb aspect morphology, on an online test).

Grimshaw, Adelstein, Bryden and Mackinnon (1998) report the case of a teenager, E.M., profoundly deaf from birth, who was first fitted with hearing aids at 15. Isolated from other deaf individuals for most of his childhood, he communicated in "homesign" with his family, and had received oral training for some months at the age of 12 9at a school for the deaf). The hearing aids allowed him to cope with conversation-level input. His verbal development over the following 4 years can be compared to Genie's: his syntactic development was equally limited, and his signing displayed a heavy reliance on semantic clues. Grimshaw et al (1998) conclude that

E.M. supports the case for a critical period for first language acquisition, particularly as his upbringing lacked the abusive characteristics of Genie's.

A recurring problem with the critical period hypothesis, across species and across behaviours, is that maturational changes in behaviour cannot always be matched to maturational biological changes. As Grimshaw et al. state in the study just referred to, echoing many other such reports: "Unfortunately, we have no data that can address the nature of the critical period" (253). However Colombo (1982) points out that critical period effects assessed biologically are more clear-cut (more reliable?) than those that can only be assessed behaviourally. For language, adequate biological - ie neurological - evidence for a critical period has not so far become available (for technical as well as ethical reasons), although the possibilities offered by new techniques such as neuroimaging are promising. What is clear is that (a) children are biologically privileged over adolescents in primary language acquisition but that (b) the maturing property of the brain that allows this is still unknown. The linguistic evidence is also incomplete - for similar reasons - and inconclusive: what the "system" or "mechanism" affected by maturation is, is unclear. The linguistic evidence points nevertheless to a "sensitive" period (Lamendella 1977, Long 1990), a weak version of the CPH which posits a universal and limited period during which the organism is more responsive to external stimuli than at other developmental stages. Rather than shutting down abruptly at puberty - as Lenneberg claimed - the sensitive period seems to start to tail off from around 7 years of age (possibly earlier – Newport and Supalla 1990 cite 4 years) and there may be different sensitive periods starting and finishing at different times for different linguistic areas, such as phonology, morphology, syntax and the lexicon (as evidenced by data from second language acquisition) (Genesee 1988). Moreover, a sensitive period may allow late - albeit incomplete - acquisition, through cerebral functional reorganisation, or the use of subroutines (Colombo 1982:270).

Hurford (1991) describes an extremely interesting and suggestive computer simulation of selective pressure on the evolution of a critical period for language acquisition. Hurford argues that (a) early (pre-reproductive stage) acquisition of language confers an evolutionary advantage and (b) the language acquisition capacity does not last throughout the lifespan, but ends, also before the reproductive stage, because of "the lack of selective pressure to acquire (more)

language (or to acquire it again) once it has been acquired" (Hurford 1991:172). The capacity to learn language is thus not "switched off" at puberty by some neurological event, but rather ceases to be needed, perhaps atrophying. Running the simulation with a population of individuals afflicted by several severe language-impairing strokes during their lifetime produced an outcome showing the language learning capacity being maintained till the end of life. Hurford (to appear) attempts a synthesis of this evolutionary hypothesis with Elman's "working memory" approach, developed in computer simulations of language learning. According to this, it is the limited capacity of working memory (ie the short-term memory system involved in the temporary processing and storage of information (Gathercole and Baddeley 1993)) in the very young child (or neural net in the case of the simulation) that allows language to be acquired: neural nets that were simply exposed to random sentences never acquired the ability to predict the ends of sentences, while when input was graduated, either by ordering the presentation of sentences or by limiting initial working memory so that the system only attended to input in short bursts, the nets would soon learn to predict. Newport's "Less is More Hypothesis" (Goldowsky and Newport 1993) also argues that the perceptual and memory limitations of the child prevent her from remembering longer chunks of speech, but thereby enable her to attend to small "bits and pieces of language", those which carry meaning. The "Less is More Hypothesis" and "Starting small" certainly provide an explanation for critical period success, and for cases like Genie's, where an adult-like working memory finds it impossible to control input in such a way as to learn from it. However, this hypothesis seems unable to account for critical period offset: if this depends on an increase in working memory, an extremely brief critical period results, at odds with empirical evidence such as Alex's. Jusczyk (1997) who has studied extensively how infants process language, proposes that "the developing memory and attentional capacities of older infants bring with them not only the ability to consider more of the same type of information simultaneously but they also provide the opportunity to consider more different kinds of information... This could be another way 'less is more' is a factor in infants' sensitivity in the sound stream of language" (Jusczyk 1997:203). However, even if the development of working memory is not the whole story, and other functions, such as attention are also implicated, if the adult working memory cannot cope with language acquisition, how to explain any adult second language learning?

2.3   Second Language Acquisition

Because of the inherent problem of the lack of empirical evidence confirming or refuting the Critical Period Hypothesis in first language acquisition, researchers have looked to second language learning, and to the experiences of child and adult learners. Perhaps the most salient feature of second language acquisition is the huge variety of outcomes of L2 learning. Nearly all children exposed for several years to a second language achieve native speaker levels of grammar and – most notably - phonology (Oyama 1976, Fathman 1975). Many adults, on the other hand, never achieve more than a "pidgin" variety, most acquire enough of it to communicate fairly effectively in most settings, but only a very small minority become proficient enough to pass as native speakers[2].

Several studies have proposed that the CPH, or the weaker sensitive period hypothesis, may help to explain this variable pattern of achievement. If successful language learning is dependent to any large degree on the age of first exposure, as clearly is the case in FLA, it can be predicted (a) that most children will reach native-like proficiency and competence; (b) that adults may never achieve native-like proficiency in a second language, or that if they do, their grammatical competence will be unlike that of native speakers or of child learners.

It is important to be precise about the area we are considering. Lenneberg's original theory, although only touching on SLA, dealt with ultimate attainment, not rate of acquisition. Many studies have shown that adults, adolescents and older children are faster and more efficient at acquiring the syntax or phonology of a second language than children (Ervin-Tripp 1974,Krashen 1982, Neufeld 1979, Snow and Hoefnagel-Hohle 1978, Singleton 1995). These studies are comparing rate of acquisition, not ultimate achievement, although Walsh and Diller (1986) argue that neurological studies provide supporting evidence for more rapid and efficient second language learning *after* childhood. It is hardly surprising that older learners (at the Piagetian stage of formal operations) should acquire language faster than younger ones. It seems on the whole however that the advantage is temporary.

Michael Long's extensive review of studies in this field (1990) concluded that maturational constraints affect second language learning in the same way as first language acquisition. It follows, according to Long, that learning the first language in childhood does not aid the learning of second and subsequent languages in later life (the "exercise hypothesis" - see below). Neither does the first language inhibit the acquisition of other languages in childhood. It also follows that adults can never attain native-like proficiency: "[n]ative-like morphology and syntax only seem to be possible for those beginning before age 15." (Long1990:280). Neurological support for the position that late second language acquisition is qualitatively different was reported by Zatorre (1989) in a critique of research into hemispheric specialization and multiple languages: although there is no reliable evidence that multilinguals make more use of the right hemisphere than monolinguals, there is a "limited but very significant body of evidence indicating that multiple languages may be slightly differently organised within the language-dominant hemisphere of a multilingual subject." (130). The data from the study of Alex (Vargha-Khadem et al 1997) however perhaps suggest that right hemisphere language dominance is less problematic, and may be more feasible, for both L1 and L2 acquisition, than was previously thought. Recently, Kim, Relkin, Lee and Hirsch (1997) reported on a functional magnetic resonance imaging (fMRI) investigation of fluent bilinguals, exposed from birth (=early) and in late childhood (=late; although the timing and nature of the initial exposure is not well defined). When late bilinguals were asked to carry out silent linguistic tasks in two languages (subjects were asked to describe a day's experiences in their heads), two separate areas of activation were observed, with the fMRI technique, in Broca's area of the brain. Broca's area (the lower back part of the frontal lobe) is generally held to be implicated in speech production, and this finding suggests that "two specific regions served each of the two languages" (171). For the early bilinguals, the centres of activation were close. However, in Wernicke's area (in the upper back part of the temporal lobe) which plays a major role in comprehension, both early and late bilinguals showed close or identical centres of activation for the two languages. Although previous research had observed spatially separate representations for first and second languages, this is the first time such functions have been localized. The authors suggest that "representations of languages in Broca's area that are developed by exposure early in life are not subsequently modified. This could necessitate the utilization of adjacent cortical areas for the second language learned as an adult." (173)

Studies of the acquisition of native-like accent by second language speakers (Asher and Garcia 1969, Oyama 1976, Fathman 1975, Scovel 1977, 1981) suggest that only those exposed to the second language before the age of 6 will achieve this. Against this position are ranged arguments by Flege (1987) based on experimental evidence of adults' abilities to both recognise and produce sounds not instantiated in their L1. The undoubted reality, that adults do maintain a foreign accent when speaking an L2 where children (although not inevitably) don't, can, he claims, be accounted for by factors of exposure, motivation and inhibition and children's more flexible learning styles (which may be features of maturation) (see also Martohardjono and Flynn 1995). Flege (1987) proposes that "although the English learner of French can detect auditorily the difference between French and English /t/, he or she will not develop a new phonetic category for French /t/ because of a general cognitive constraint" (Flege 1987:170), viz. prior L1 knowledge giving rise to assumptions about the allophonic character of French /t/ and English /t/. However, if such psychosocial and individual variables as exposure and learning style are the main factors in achievement, one would expect much more variability than has been attested amongst children, and – concomitantly – more success amongst adults. But even for adults attitudinal factors are not always significant in L2 performance (Oyama 1978, Patkowski 1980, Johnson and Newport 1989). There is considerable evidence of the greater ability of older children, adolescents and adults to make use of formal, intensive training (Thogmartin 1982, Ervin-Tripp 1974, Ekstrand 1976) but it is equally true that in non-experimental settings and in naturalistic acquisition adults tend to show significantly greater foreign accents, independently of their grammatical proficiency (Patkowski 1990). There is evidence that phonetic perception abilities do not deteriorate with age (Wode 1994) as phonological production seems to. This asymmetry suggests a critical period for speech production, but not for perception. Multiple critical periods for language are hardly parsimonious theoretically, and seem intuitively improbable: if they can be demonstrated however, the notion of critical period itself is weakened.

Findings like these do pose the question of different critical periods for syntax and for phonology. Walsh and Diller (1986) hypothesize that different domains may be subserved by different types of brain cell: phonology is subserved by pyramidal cells which cease to develop at the age of 6 or 7, while the stellate cells responsible for higher order domains like morphology and syntax

continue to develop well into young adulthood. However, there is evidence that some adult learners at least can acquire L2 phonology (see Young-Scholten 1994).

For syntax, Oyama (1978), Patkowski (1980), and Johnson and Newport (1989, 1991) investigated ultimate attainment and found that performance in grammaticality judgement tests, on a variety of language specific and UG-constrained features of English, began to decline amongst learners first exposed to the second language after 7, although Patkowski set the optimum age for second language acquisition at 12-15. None of these early investigations found that language learning ability remains at the same level throughout childhood and early adolescence, nor that there is a sudden and catastrophic loss of the ability, at any age.

Patkowski (1980) proposed a modification of the original CPH. He notes that Lenneberg (1967) considered that the existence of successful adult second language learners

> does not trouble our basic hypothesis on age limitations because we may assume that the cerebral organization for language learning as such has taken place during childhood, and since natural languages tend to resemble one another in many fundamental aspects, the matrix for language skills is present. (176)

Patkowski (1980) distinguishes between a critical period in FLA and a sensitive period in SLA In the latter, second language learning during the critical period may result in complete acquisition (given the right socio-psychological conditions) although this outcome is not inevitable. Conversely, second language learning after puberty is feasible but never results in native-like proficiency, although "extremely high, quasi-native levels can[...] be attained in one or more areas"(466) for example in syntax, but not phonology. His study, of 67 immigrants resident in the USA for at least 5 years, concluded that even highly educated learners living in "the best" conditions for successful acquisition do not attain native-like levels of English in all areas unless they start learning as children. He had native judges rate the syntax of transcripts of interviews. Their assessments show a normal distribution for the adult learners but a right-skewed curve for the pre-puberty learners. A second experiment using a grammaticality judgement test produced similar results. Only age of first exposure and *not* length of stay or amount of formal instruction or of informal exposure correlated significantly with the syntactic ratings. Patkowski concludes

that, while socio-cultural and attitudinal factors clearly affect both rates of learning and ultimate achievement, the theory "most compatible" with his results is that which proposes an innate, genetically-determined mechanism for language acquisition (467).

The most reverberant investigations of the critical period hypothesis in the acquisition of the syntax of a second language have been undertaken by Elissa Newport and Jacqueline Johnson. Their two studies (1989,1991) of Chinese and Korean learners of English added powerful support to the position of Patkowski and Oyama (and Coppieters 1987 - see below, chapter 3) by demonstrating that learners who are not exposed to a second language till adulthood are less likely to achieve a high level of grammatical competence than children, although they are not barred from a partial knowledge. Both studies have attracted a great deal of attention.

Their first study (henceforth J&N 1989) examined two hypotheses: the "exercise hypothesis" predicts that so long as a first language is acquired within the critical period, there will be no prohibition on the successful acquisition of second and subsequent languages in later life. The "maturational state hypothesis" limits all language learning capacity to the critical period. Their findings tend to support the second hypothesis. They tested 46 Chinese and Korean learners of English who had lived in the US for 10 years on average (range 4-26 years) on 12 basic structures of English (including plurals, word order and auxiliaries) with an aurally presented Grammaticality Judgement test. Half of their subjects had arrived as children (before the age of 15), half as adults (after 17). The test results show that the subjects who arrived before the age of 7 could match native speaker levels but "[i]n each of the remaining groups [8-10, 11-15 and 17-39], as age of arrival increased, performance became significantly poorer". In fact, mean scores dropped from 269 (out of 276 possible) to 210: 99.6% for natives and the 3-7 group; 92% for the 8-10 group; 84% for the 11-15 group; 76% for the 17-39 group. The decline in ultimate performance began at around 8 years and fell steeply and steadily into adulthood, as J&N's Figure 1 demonstrates:
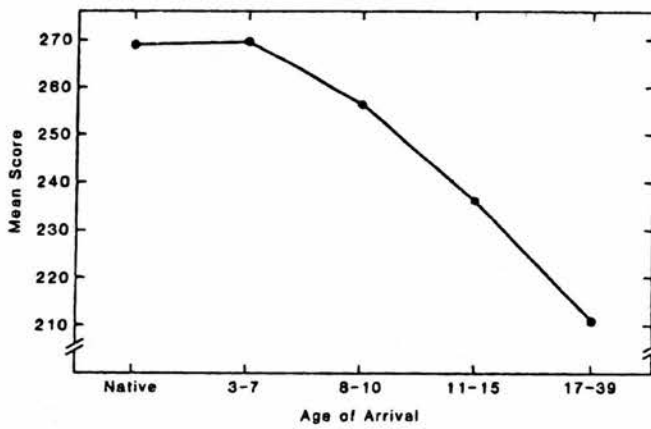
Figure 2.1. Age of arrival and total score correct on test of English grammar (from Johnson and Newport 1989:79)

The data show no sudden drop in performance at puberty, but they claim that the correlation between test performance and age of arrival ceases to hold thereafter: "however there are large individual variations in ultimate ability in the language, within the lowered range of performance" (J&N1989:81):
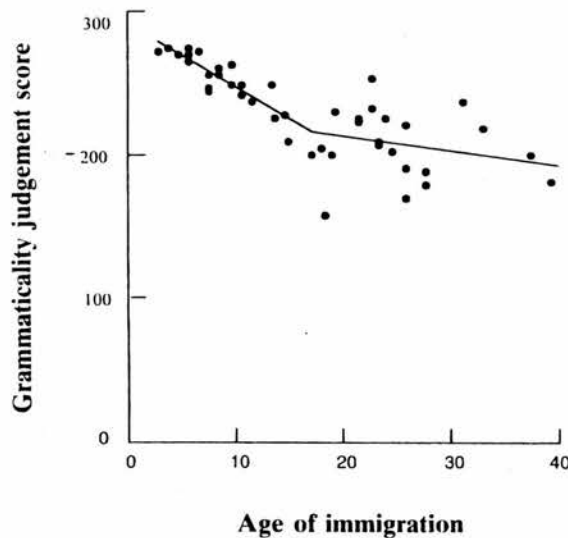
Figure 2.2: Results of grammaticality judgements by age of arrival, adapted from Newport and Johnson (1989) (Bialystok 1997:122)

Notice that one learner who arrived at the age of 23 scored well within the range for 3-7 year olds, although for the rest there was almost no overlap with the child arrivers.

J&N also tested for affective and socio-cultural factors and found that only when combined with age of arrival did lack of "self-consciousness" and "identification with America" have a significant effect on test performance. This result is consistent with Oyama (1978).

J&N's findings support a weak version of Lenneberg's hypothesis: a steady decline in language learning ability from the age of around 7 to adolescence, and the continued possibility for adults

to learn second languages, although with a far broader range of final levels of attainment than is found amongst child learners. They suggest that this variability raises the question as to whether adult language learning is "controlled by a different set of variables" (97), presumably including instruction and the quality and quantity of L2 input. This raises the question of the validity of comparing L1 subjects with any other than naturalistic L2 learners. J&N's findings are however also consistent, as Hurford (1991) has pointed out, with an "interference hypothesis: second language learning is (to some extent) inhibited by prior attainment in a first language" (Hurford 1991:163).

J&N therefore posit a sensitive period, which differs from a critical period in that the age-related decline in ultimate achievement, as measured by their test results is gradual rather than abrupt. Child learners with an age of arrival as young as 4-7 are beginning to display non-native intuitions about English, while amongst adult learners "there appears to be some residual, though greatly diminished, ability to learn" (1991: 255). Patkowski (1980) also favours "sensitive" rather than "critical", and the former epithet seems also appropriate for late first language learning of ASL (Newport and Supalla 1990, Mayberry and Eichen 1991). Mack (1984) looked at aspects of phonological and syntactic knowledge of monolinguals and early bilinguals (age of arrival before the age of 8) and found differences attributable either to the "intrusion" of the other language or to a different organisation of knowledge of *both* languages in the bilingual brain, compared to the monolingual brain. Hyltenstam (1992) compared the written production of active bilinguals in Finnish and Swedish who acquired their second language before and after the age of 6. Although the error rate was very low for both groups, there was a clear difference between them, leading Hyltenstam to speculate that fossilization may set in much earlier than other writers, such as Long (1990), have suggested. There thus seems to be considerable evidence that non-native intuitions about language particular features of the L2 start to appear in quite early SLA. However, a problem with both the notion of a sensitive period, and with the empirical evidence is highlighted by Snow (1987) who notes that Patkowski's sensitive period seems to start around 12, Johnson and Newport's at 7-8, and that "no particularly salient change in brain functions" occurs at the earlier age.

J&N (1991) continue the investigation by asking whether universal features of language - those properly belonging to Universal Grammar - are as constrained by the critical period as language specific features. They tested 23 adult Chinese learners of English on their knowledge of universal and parameterised rules of Subjacency (a constraint on question and subordinate clause formation) with an aural judgement test. Subjects were again selected on the basis of a minimum length of residence (5 years) and not through a proficiency test, although one subject was eliminated because he was not performing above chance on wh-movement in simple sentences. The test procedure was similar to J&N 1989: 180 test items were presented on audio tape and rated dichotomously. The adult learners scored significantly below the native speaker controls, although above chance on two types of constraint violation. Results on the language-specific items (questions violating the inversion rule) were rather better.
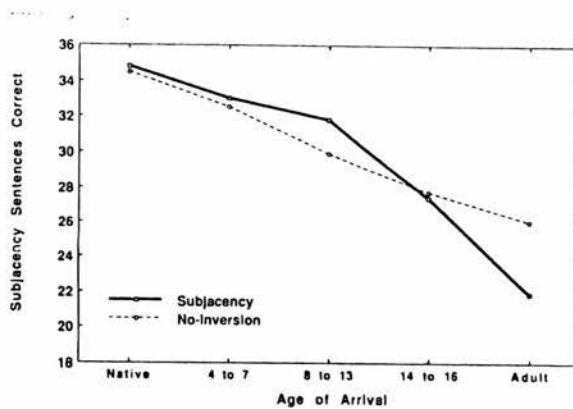
Figure 2.3. Age of arrival and total correct score on Subjacency and no-inversion test items (J&N 1991: 248)

J&N conclude that principles and parameters of Universal Grammar are not privileged in second language learning over language specific rules. Similar results were obtained from a study of child learners reported in the same paper. The patterns of responses did not match any non-English settings of the Subjacency parameter such as the Russian (more restrictive) or the Italian (less restrictive) and so it would appear that their knowledge of Subjacency was not constrained by Universal Grammar at all - was not a *possible* grammar. Subjects may however have performed better on the no-inversion violations because all had been taught English in high school or college

(mean of 6.8 years of instruction), and this rule is one of the earliest taught in all English courses (regardless of the pedagogic approach), and is also highly salient in naturalistic input. It is also frequently overlearned, appearing in embedded questions for example. The Subjacency condition, on the other hand, is not taught explicitly, is extremely rare in the input, and is late-acquired in first languages. Given that there is no indication of their proficiency relative to native speakers, it is perhaps unsurprising (a) that the subjects performed poorly on Subjacency compared to native speakers, and (b) that results on the Subjacency violations were worse than those on no-inversion.

While Johnson and Newport's studies clearly lend strong support to a maturational account of the variable attainment of adult learners, the question of adult achievement can be posed rather differently: is it *possible* for adult learners to achieve native or near-native levels of proficiency? J&N (1991) conclude that "the subjacency principle is not fully accessible to the mature learner for the learning of a second language" (243) and suggest therefore that "the older learner has both a weakened or diminished set of universal constraints on human languages in general" (256). The figures reproduced above show clearly that the Subjacency principle is *partially* accessible and J&N do not offer a theoretical rationale for the "weakening" of UG constraints beyond hypothesizing a gradual shift from the language acquisition device which enables the learning of a first language to (unspecified) "entirely different, and less constrained, learning mechanisms"(256).[3]

Both these studies made a considerable impact on the critical period debate. They have attracted some criticism (Long 1990, Birdsong 1992, Kellerman 1995, White and Genesee 1996, Murphy 1997, Bialystok 1997), and have also been replicated, with rather different results. Two critical analyses are considered here. Bialystok in a recent critique (1997) challenges the conclusion that the differential performance of older and younger learners is caused by the maturational factors associated with the critical period. She suggests - not very plausibly - that the lack of variability found in the test results of the latter could be due to (a) their shared educational background in US schools, which the adult arrivers lacked, and (b) their longer years of residence[4]. As for the native-like performance of the pre-7 year-olds, Bialystok speculates that English, not Korean or Chinese, was their dominant language from the start. Although J&N clearly state that they

excluded (child) subjects for whom English was the main language spoken in the home, an important issue is raised here: few of the major studies in this field test subjects for L1 proficiency or even investigate the extent of L1 use at the time of the study. Hyltenstam (1992), referred to above, is a rare counterexample. In the case of child learners particularly, it may be quite inaccurate to term them L2 speakers or bilinguals, and the linguistic competence being tapped may - in reality - be more monolingual than bilingual.

Bialystok (1997) contends further that J&N's data reveals a continuing age effect into adulthood, suggesting not a maturational effect but one that can be explained by learning styles. Thus children tend to create new categories for new phonological or syntactic features (accommodational strategy) while adults are more likely to extend existing categories (assimilative strategy). She argues that exceptional adult learning can be explained better by cognitive and affective - and environmental - factors than simply maturation.

Birdsong and Molis (1997) combined the test data for all subjects in J&N 1989 on a single scatterplot (Figure 4) and show that an age of arrival effect continues into adulthood, thereby immediately calling into question the equation: age effect = critical period. Secondly, a replication study with Spanish learners of English shows a significant effect of age *only* amongst adult, not amongst child learners (r= -.22 for early arrivers, -.69 for late arrivers). Their data strongly indicates a critical, not a sensitive, period effect: test performance starts to decline with adolescent arrivers, not earlier (apart from three unusually low-scoring 8-10 year-old arrivers).

Birdsong's (1992) replication of Coppieters' (1987) study of ultimate attainment will be discussed in detail in chapter 3. Here, it is appropriate to report his findings in relation to age of arrival effects. Birdsong's subjects were 20 near-native speakers of French whose L1 was English, and whose exposure to French began after puberty. The main predictive variable of results for the Grammaticality Judgement test he administered was age of arrival in adulthood, again *contra* J&N (1989; 1991). Thus subjects who arrived in France before their early 30s showed a pattern of deviation from NS norms similar to the native speakers themselves, while those who arrived after the age of 35 were more likely to diverge markedly in their intuitions, suggesting either a "critical period" with an offset time near middle age, or some other age-related factor. Another study of

exceptional adult language learners was carried out by Ioup, Boustagui, El Tigi and Moselle (1994). This study is also discussed in depth in chapter 3. Both the learners they describe started to acquire their L2, Egyptian Arabic, well after puberty, and both attained near-native levels in both production and metalinguistic tests. The main subject was Julie, who acquired her L2 naturalistically. Ioup et al contend that this clear breach of maturational constraints was brought about by two exceptional factors: one behavioural, one biological. First, Julie was unusual in the extent to which she paid conscious attention to form. For example, although she did not attend Arabic classes, she kept a notebook from her first days in Egypt in which she noted down aspects of the language; she also asked questions. In other words, she was self-taught (rather than untaught), and coped with language learning in a very different way from what is known about naturalistic learners.

Ioup et al (1994) speculate that Julie has an innate talent for language learning, which itself is "hypothesized to be associated with unusual brain organisation where a greater proportion of cortex is devoted to language ... allow[ing] the learner to be more cognitively flexible in processing L2 input and ultimately organizing it into a system."(92). Ioup et al (1994) conclude that either talented learners manage in some way to retain access to the language acquisition system or device, postulated for child L1 learning, or they tap into (perhaps more deeply into) a separate neuropsychological function, again inaccessible to most adult learners.

While it seems plausible that language-specific structures should be acquired through the gifted learner's superior grasp of form-function relations, associative memory, and mastery of new codes, how they are able to access UG more completely and more successfully than the majority, as Ioup et al suggest, is less clear. It may be that the advantage the gifted language learner has over the rest is more efficient ways of "getting past" their L1 knowledge so that they access UG directly. In other words, the language acquisition function is not switched off, or dimmed, at adolescence, but merely becomes more difficult to operate: exceptional learners are those who still can operate the device; alternatively, they are more skilled at general cognitive operations and manage to mimic the child's method in some – as yet unspecified - way. More studies utilising neuroimaging techniques may in future shed light on this issue.

The studies reviewed above focus on the differences in ultimate attainment between child and adult learners. The claim that child learning is privileged over adult learning is irrefutable. The reasons for this, although apparently biological rather than psychological, have still not been established, although the plasticity of some functions of the brain seem clearly implicated. The intertwined issues of the end of the critical period and the language learning capacities of adults have received less attention from researchers. As we have seen, research findings on the effect of age of arrival on adult learners are contradictory. If maturation is involved, an investigation of adult learners should find - as Johnson and Newport did - that age of arrival does not correlate with knowledge of the L2 grammar. Whatever the facts of maturing neurological processes during the sensitive or critical period turn out to be, there can be no maturational effects after adolescence. If there is an age effect in adulthood, as Birdsong (1992;1997) finds, then not only the critical period is at work: the age effect in adulthood represents something other than maturation. This alone justifies further empirical investigation. In addition, the story of language learning abilities through adulthood deserves closer inspection. The next section reviews work in this area done in both first and second languages.

2.4   Language Abilities in Later Life

There is relatively little research evidence on what happens to language abilities in later life. In surveying the literature here, three questions are posed:

(1) Is there evidence of biological changes during adulthood that affect language abilities?
(2) If so, are these changes in grammatical competence or in memory and processing mechanisms?
(3) Is it possible to distinguish between language acquisition capacities and language maintenance capacities within the language function?

Leaving aside for the moment social and psychological factors which undoubtedly have an impact on language learning success in adulthood, what neurological changes or conditions could account

for a continuing age effect? It is instructive first to survey the literature on L1 maintenance across the lifespan.

Research into the language abilities in later adulthood, both monolingual and bilingual, strongly suggests that deterioration in short-term memory is the most significant neurological change while language knowledge, or competence, appears to be unimpaired by ageing. Kemper (1987,1989) looked at the oral and written productions of older adults in their first language (English) and found that they used significantly fewer complex sentences after the age of 40. She attributed these language deficits to a decline in working memory: memory span correlated positively with the incidence of left-branching sentences. This applied to processing as well, while adults of all ages were able to process right-branching sentences without difficulty. Other studies have found similar effects: older people, especially those over 75, take longer to process sentences and make more errors with syntactically complex sentences, because of short-term memory decrement. Some studies have found such effects in subjects under 40 (Emery 1986). Pye, Cheung and Kemper (1992) looked not at performance but at grammaticality judgements of three types of sentence - anaphoric, wh- questions and left- and right- branching - in shorter and longer formats in order to establish "whether elderly adults merely lose their capacity to process sentences or whether they undergo irreversible changes to the structure of their grammatical system"(356). Their study confirmed that sentence length and complexity has a main effect on elderly people's ability to distinguish between grammatical and ungrammatical sentences. While they had no difficulty in detecting rule and constraint violations (such as Complex NP, and specified subject and tensed-S constraints), respondents tended to mark (on a 7-point scale) as ungrammatical grammatical sentences which were long or complex, notably long sentences with anaphors, and doubly-embedded wh-island sentences whether short or long (Pye et al:369). There was also some age-related decline in their ability to recognise short simple grammatical sentences. As the only variable that showed a main effect was short-term memory, and not years in education or vocabulary size, it is suggested that "adults' linguistic competence remains intact across the life-span...[and that]...judgements of sentence grammaticality are subject to performance limitations on the processing of long or complex sentences"(370). In other words, old people do not undergo some reverse process of first language acquisition during which their language knowledge or competence becomes simplified; instead, performance (or "control" to use the term adopted by

Bialystok and Sharwood Smith (1986)) has become eroded by nonlinguistic factors. Kemper (1998) tested healthy older adults (average age 68.5) and adults with probable (ie not severe) Alzheimer's Disease, and was able to show that even the latter demonstrated intact syntactic knowledge.

Unsurprisingly if the view that performance factors like short-term memory are distorting GJ test results holds up, choice of test procedure may be a confound. Pye et al gave their elderly subjects the test items in booklet form, so that they could judge them in their own time, at home. Emery (1986) on the other hand appears (test modality is not specified) to have administered her tests orally (moreover, the semantically complex sentences appear extremely difficult): recall that she found a short-term memory effect starting with subjects in their 30s. A study by Feier and Gerstman (1980) which compared differences in the comprehension of relativised sentences between 4 age-groups across the adult life-span (52-58, 63-69, 74-80, with a control group of 18-25 year-olds) produced similar results. While they found that error-rate jumped dramatically for the 74-80 year-olds, performance declined stepwise through the younger groups. On the other hand, only half of the performances were markedly inferior to those of the younger groups (727). However, intelligence, as measured by the WAIS subtest of vocabulary size, and years in formal education alongside memory span did predict variation within age-groups. Interestingly, sentence type did not, contrary to expectation, interact with age. There are serious flaws in the experimental design of this study: a toy moving activity was used with the elderly respondents and results show a very low error rate and very high intra-group variability (half the responses of the 70 year-olds were indistinguishable from those of younger respondents). It is hard to resist the thought that the nature of the task itself may have been responsible for such results, as the authors themselves admit (Feier and Gerstmann 1980:72).

A recent study confirms the findings of Pye et al (1992). Zurif, Swinney, Prather, Wingfield and Brownell (1995) in an investigation of the operation of working memory during the processing of filler-gap dependencies (ie complex subject and object relative clauses) predicted that age-related memory deficits would impact on the semantic interpretation of complex sentences, but would have no impact on on-line syntactic processing. They tested monolingual people aged between 60 and 76 years. In fact, results showed that the longer the distance between antecedent

and gap the greater the difficulty. However, within the reduced memory storage space, their elderly subjects had no problems with syntactic complexity. Other studies (Gunter, Jackson and Mulder 1995, Brebion, Ehrlich and Tardieu 1995, Kemper 1997) also suggest that in old age and even middle age, working memory deficits are beginning to affect language tasks, although individual differences may also be at work. Wode (1994) points out that age does not apparently stop the acquisition of new lexis, or new phonology in either L1 or L2: lexical and phonological changes in speech communities tend to be introduced by adults, not by children (see studies by labov in Trudgill 1983).

Conversely, a recent study by Bloom, Mullins and Paternostro (1996) proposes that "a subtle decline in syntactic processing begins before the age of 70" (112). This study tested three age groups with a reading task and a controlled writing task on the use of adverbial conjuncts. The young and middle aged groups (20-25 and 50-55 respectively) produced similar results on the former task, but on the writing task there were indications that at 50 the ability to use terms like "therefore", "furthermore" and "however" correctly is beginning to go. The nature of the task - somewhere between spontaneous production and a GJ test - makes it difficult to compare this study with Pye et al (1992) and their ilk.

A brief survey of the literature on neurological changes and cerebral insults can help to illuminate the later development of the language function. Brain damage caused by illness or by severe injury to the head can result in language loss (aphasia). Later, patients may recover their speech, wholly or partly. Aphasia studies however (Paradis 1977) suggest that when a bilingual aphasic patient is unable to recall one of her languages, it may not be irretrievably "lost", merely irretrievable: aphasics may understand a language without being able to speak it, and many patients do recover their language(s) after a time. Moreover, recovery of a language is not straightforwardly determined by the way it was acquired. A statistical survey of clinical data on bilingual aphasics who did not retrieve their first language (Obler and Mahecha 1991) found that age, degree of bilingualism and type of bilingualism (compound, co-ordinate or compound-co-ordinate[5]) were *not* significant in predicting patterns of recovery. They propose that "rather than factors of bilingual history *per se*, it would appear to be brain-based factors that determine recovery from aphasia in the L1. In particular it would seem that individuals with an unusual brain organization

(*viz.* left-handed individuals) are most likely to lose their L1"(63). It should be noted that findings from aphasia studies must be treated with some caution, as the primary data they use is unlikely to be representative - it is the notable and exceptional cases that are written up.

Most of the linguistic literature on bilingual or L2 speakers in old age is concerned with loss. Parallels between patterns of loss and patterns of acquisition have been detected, in particular in the way that both processes involve syntactic simplification, transfer and interference (Sharwood Smith 1989). There is some evidence - though often anecdotal and impressionistic - that the capacity to maintain two or more languages declines with age.

In some countries, as De Bot and Clyne (1989) relate, an apparent retreat from the second language among elderly migrants has encouraged policy decisions to open mother-tongue old people's homes, day centres etc. Their investigation into language loss among elderly migrants in Australia reported that although observations showed that elderly German-speaking subjects increased code-switching into the L1 and became less fluent and less accurate, when older Dutch speakers were asked to evaluate their performance in English, there was no difference in the average ages of those who reported a decrease, in old age, in proficiency and those who reported an increase. However differences in exposure, which in this longitudinal study could not be calculated, may mean that some of the respondents never achieved the "critical threshold" of language acquisition, ie the point at which knowledge becomes immune to interference or decay (Neisser 1984). The language reversion hypothesis (see De Bot and Clyne 1989) proposes that as migrants age, both reversion to the L1 and attrition of the L2 become evident. They evince increased code-switching - both code-mixing, when lexical items from the dominant language are inserted into utterances in the weaker language, and code-changing, where whole phrases or sentences are borrowed; decreased fluency and accuracy in the L2; a more marked "foreign" accent and a preference for the L1 in more intimate areas of daily life such as religion. De Bot and Clyne (1989) suggest that the incidence of such features is not in itself proof of attrition or of actual language loss and may be attributed to social and psychological factors such as the desire to re-identify more closely with the L1 culture or to return to one's roots in retirement, or when children left home – the pressure to use the L2 in many domains being reduced

To sum up, monolingual language competence and competence in a second language (if that language has been learned to a certain level) is not lost in old age, and is rarely completely lost even in cases of aphasia. Both the psycholinguistic and the neurological evidence suggest that the language function of the brain remains resilient and flexible in adulthood, and even in old age. Performance on the other hand both in the first and in the second language may be impaired in older people by processing difficulties, caused by the deterioration of short-term memory as well as by psychosocial factors such as those identified by De Bot and Clyne (1989). We now look at other possible effects of short-term memory deficits on language learning.

2.5   Memory and Language Learning

The study of memory and learning is a huge field and it is not the purpose of this investigation to attempt to review it. Nevertheless, it is relevant here to refer to some of the work that has been done on the relation between language learning and short-term or working memory at different life stages. Gathercole and Baddeley (1993) provide an excellent summary of this research. Using Baddeley and Hitch's (1974) Working Memory model, which comprises a central executive and two "slave" systems, the phonological loop and the visuo-spatial sketchpad, they show that all three components become more efficient with age through childhood and into adulthood: there is no critical period for short-term memory. While the total neural space available for processing remains constant throughout development, the more efficient and automatic processing becomes the more space can be used for storage, thus increasing memory span (Case, Kurland and Goldberg 1982). Hence, the vocabulary learning skills of 11 year-olds are more effective than those of 6 year olds, although there is wide variability. Gathercole and Baddeley (1993) attribute this variance between individuals not to environmental factors, but specifically to the effectiveness of the phonological loop (a structure capable of holding and rehearsing a small amount of speech-based information while the central executive is freed for other mental tasks). Gathercole (1998) in a recent review concludes that from infancy to 7 there are substantial qualitative changes in phonological memory, with only gradual improvement thereafter (to early adolescence). In the domain of sentence processing and comprehension, getting subjects to inhibit phonological

memory by the technique of articulatory suppression significantly reduces error detection in both long and syntactically complex sentences (Gathercole and Baddeley 1993), and on the learning of foreign language words (Papagno, Valentine and Baddeley 1991), although in a study of patients with a stroke-damaged phonological loop system subjects were able to acquire new words from a visual presentation (Gathercole and Baddeley 1993: 44). Miyake, Carpenter and Just (1994) provide supporting evidence from an experiment which involved reducing the working memory capacities of normal adults; they found that this severely impaired comprehension, as with aphasics.

Gathercole and Baddeley (1993) further argue from, *inter alia*, studies of language impaired children that up to the age of 5 the direction of causality is from short-term memory to vocabulary (see also Isaki and Plante 1997); after 5 reading ability takes over as a main factor (although prior learning may be an intervening variable). In post-5 second language learning, however, working memory again becomes a main predictor of proficiency: Service (1992) found a significant correlation between pre-learning repetition scores (ie phonological memory) and school grades in English two and a half years later for Finnish children who began to learn English at 9 years. These scores did not correlate with grades in arithmetic, suggesting - yet again - the independence of the language faculty from general cognitive abilities (see also Cheung 1996).

Experimental studies have shown that adults learning foreign words also make use of the phonological loop component of working memory when they are unable to make associations with L1 words (Baddeley, Papagno, Vallar 1988, Papagno et al 1991). Harrington (1992) found a strong correlation between working memory and adult L2 proficiency, although the causal connection here was less convincing (see also Papagno and Vallar 1995). Brown and Hulme (1992) show that *long-term* phonological memory is also implicated in L2 success (Baddeley, 1998). On the other hand Berquist (1997) found an opposite trend: in a study of young adult French learners of English where L1 and L2 word spans and reading spans were tested and compared with TOEIC (proficiency) scores, positive correlations were found between reading proficiency and L1 reading span and between L1 and L2 reading spans (ie working memory + processing); Berquist however identifies the direction of causality as processing → working memory, rather than vice versa (L1 word span does not correlate strongly with either L2 reading

span or TOEIC). Hence, he argues, working memory as revealed in L1 tests does not predict L2 achievement, rather the predictor is energy of processing. "The variation in W[orking] M[emory] capacity among individuals in L2 does not appear to be directly linked to what could be explained in L1 as genetic differences." (Berquist 1997:5). However, regardless of which component of memory processing is responsible, short-term or working memory in the L2 is less efficient than in the L1.

Studies of short- term memory suggest a slight decline in short-term memory in sentence processing due to age (Kausler 1992, Gunter et al 1995). Age related deficits in word span are found by Kynette, Kemper, Norman and Cheung (1990 - see in Kausler 1992), and again linked to the slowing of rate of articulation.

If short-term memory is closely involved with language learning, and becomes more effective with age through childhood, this could at least partially account for the superior *rate* of learning of older children and adolescents, and perhaps of young adults. However, the studies reviewed in the last section, which relate processing deficits in older adults to decrements in short-term memory suggest that the latter does not affect language competence. As regards ultimate attainment in SLA, the following hypotheses are possible:

(1) Short-term memory capacity has no bearing on ultimate attainment amongst young children, whose short-term memory is less efficient than that of older individuals (perhaps reaching adult levels at adolescence) (see Elman 1993);

(2) Short-term memory is highly variable amongst adults, and is strongly implicated in near-native proficiency achieved by post-critical period learners;

(3) As short-term memory capacity correlates negatively with age, older adults are less likely to achieve near-native proficiency than younger adults.

(4) If memory capacity, including processing capacity, is a strong predictor of adult L2 proficiency, memory training could improve the success rate of adult learners.

While none of this evidence bears directly on the question of whether or not late-acquired second languages can match the proficiency, in knowledge or in performance, of languages acquired in

childhood, it does raise the question: if language competence remains robust throughout the lifespan, does the *capacity* for successful language acquisition in later life remain? Alternatively do the well-attested age-related deficits in processing capability close the route to native-like achievement for most adult learners, although not for all?

*

---

[1] Teuber (1975) proposes a "crowding hypothesis" according to which nonverbal functions are sacrificed to verbal functions when the right hemisphere is obliged to assume responsibility for language in the developing child, as language communication skills have higher priority in everyday life.

2.Selinker (1972) proposed that perhaps 5% of adult learners achieve native-speaker competence. "It follows from this assumption that 5% go through very different psycholinguistic processes than do most language learners" (212). Alternatively this minority may simply follow the same course as the rest, only more efficiently.

3.Johnson and Newport (1991) quote Otsu (1981) who found that children obeyed the Subjacency condition as soon as they acquired wh-movement. They found that their subjects' performance on the grammatical control sentences -eg "What did the teacher know that Janet liked" was not much better than on the Subjacency violation items like "What did the teacher know the fact that Janet liked?". This strongly suggests that many of their subjects had not yet acquired wh-movement., in which case the argument against UG access would seem not able to be addressed.

4.Length of residence has generally resisted implication in GJ test results of L2 speakers who have passed the "threshold" of 3 or 5 years of residence in the L2 community (Patkowski 1980, J&N 1989, J&N 1991, White and Genesee 1996)

5.'Studies by Weinreich (1953) and Ervin and Osgood (1954) tried to establish typologies of bilingualism, based on how the two lexicons are stored in the brain: "co-ordinative bilinguals" store different lexical items (eg 'book' and 'livre') in separate semantic categories; "compound bilinguals" store the two lexical items in the same meaning category. These psycholinguistic distinctions may relate to patterns of use (separate domains vs continual switching). Grosjean (1982) however points out that individuals may shift from one type to another throughout their lives, and that in any case they are not mutually exclusive. Moreover, evidence for the existence in the bilingual brain for one or two lexicons is equally impressive (Grosjean 1982:245).

CHAPTER 3

THE PUZZLE OF ULTIMATE ATTAINMENT

3.1   What is Ultimate Attainment?

3.2   Different Kinds of Competence?

3.3   Operationalising Ultimate Attainment

3.4   Ultimate Attainment Studies

3.5   UG Availability and Native-like Competence

3.6   Age Effects and Ultimate Attainment

3.1   What is Ultimate Attainment?

We have seen - in the last chapter - that adult learning of a second language results in a wide variety of achievement. The Critical Period Hypothesis itself entails non-native ultimate attainment by adult learners but, in part spurred on by the challenge of the CPH, researchers have investigated cases of near-native attainment by such learners. This chapter discusses five recent studies, all of which provide evidence that adult learners *can* achieve native-like proficiency, and even, in some areas of grammar at least, native-like competence in the second language. By the term "competence" within the framework of a UG approach, we intend an abstracted grammatical knowledge which it is possible to detect, albeit imperfectly, by means of investigative techniques such as grammaticality judgement testing. The studies described here all make use of these techniques to delineate the competence of L2 learners.

The final state of first language acquisition is described by Chomsky as a "steady state" (1986). This final state of L1 knowledge varies little from speaker to speaker, as far as can be verified empirically. L2 speakers, even from the same language background, do not achieve a steady state like this. Cook (1991) has described the end-state of second language learning as "terminal knowledge" which varies from one speaker to another, even as regards core areas of the L2 grammar. Cook suggests that, quite apart from the question of access to Universal Grammar discussed in Chapter One above, input is manipulated in classroom instruction in so many different ways that it is impossible to generalise about it in the way that we can generalise about L1 input to children. Moreover metalinguistic explanation is available to adult learners and even to older child learners, which is not the case for a 2 year-old child. Therefore, Cook maintains, "the uniformity requirement cannot be met"(109).

There is however one respect in which L2 learners' knowledge does meet the "uniformity requirement": they have all reached the steady state for L1 knowledge by the time they are exposed to the L2. The Full Transfer/Full Access model of Schwartz and Sprouse (1996) states that "all the principles and parameter values as instantiated in the L1 grammar immediately carry over as the initial state of a new grammatical system on first exposure to input from the target language (TL)" (41). The way that the learner progresses is by restructuring the starting grammar, in line with UG principles, as representations fail to be

assigned to L2 input. This may happen more or less rapidly, and will be determined partly by learnability considerations, but the ingredients of the final state will always be: UG (hence Full Access), the L1 grammar and L2 input. Studies of naturalistic learning of second languages strongly suggest that learners from a particular L1 will follow a roughly similar path of development, which will be different from that taken by learners from another L1. As for such learners input is the same - not manipulated, little or no negative evidence - differences between L2 and TL grammars and between different L2 grammars can only derive from the L2 initial state/L1 final state. In addition, this model copes neatly with the problem of fossilization in L2: "convergence on the TL grammar is not guaranteed ... either the data needed to force restructuring simply do not exist (eg negative data ...) or the positive data needed are highly obscure, being very complex and/or rare". (42)

The Minimal Trees approach (Vainikka and Young-Scholten 1996) also proposes transfer, but only of lexical projections, not of functional projections which emerge from the interaction of L2 input with X Theory. Minimal Trees can explain the successive appearance of functional projections and evidence of similar acquisition orders in the interlanguage of learners from different language backgrounds. However it is less apt than FT/FA to explain fossilization at different stages of acquisition or interlanguage features which clearly show evidence of L1 transfer (eg adverbial word order). Vainikka and Young-Scholten (1996) argue that L2A parallels child L1A, but this does not seem to cope satisfactorily with variable final states for L2A.

Using production data from the naturalistic acquisition of German by a Turkish speaker, Cevdet, Schwartz and Sprouse (1996) show how he achieves German word order (V2) for some utterances only, thus leading to fossilization. The salience of V2 in the German input forces early restructuring via an option available in Turkish (verb-raising to C) with non-subject elements being fronted by adjunction to CP (also derived from the L1). This results in the order $XSV_{[+F]}$ (eg "spater der Charlie wollte zum Gefangnishaus" - "later Charlie wanted to go to the prison"). Subsequently, pronominal subjects are moved to postverbal position, with nonpronominal subjects being moved later (eg "das hat eine andere Frau gesehen" - "another woman saw that"), resulting in target-like V2. However adjunction to CP cannot, according to Schwartz and Sprouse, be unlearned by positive evidence alone. They predict that Cevdet will never learn that $XV_{[+F]}$ S is obligatory, not optional, and that he will fossilize on the word order $XSV_{[+F]}$. Although he will hear only V2 utterances in matrix clauses, he

will never hear utterances indicating that V3 is ungrammatical.

The Full Access/Full Transfer model is useful in both characterising the difference between L1A and L2A end states and in explaining why adult (and child) learners can fossilize at non-native levels, even though the Interlanguage is constrained by UG. However, we have not yet determined (a) whether it is possible for an L2 grammar to match the native grammar exactly, or (b) whether fossilization, by far the most common outcome, represents an incomplete or a divergent version of the target grammar. By "incomplete" we mean having a narrower grammar than native speakers, so that, for example, optional constructions are identified as obligatory, while "divergent" signifies a non-target or interlanguage variant, which is still constrained by UG. As Sorace (1993), who has developed this paradigm, points out:

> The incomplete grammar, lacking a given L2 property *P*, would lead to random, inconsistent, in short indeterminate judgements about *P*, whereas the divergent grammar, since it incorporates an alternative representation of *P*, would lead to determinate judgements that are consistently different from native judgements. (24)

Knowledge of UG principles and parametric values may not necessarily entail obedience to them in performance, even in metalinguistic tasks. Grimshaw and Rosen (1990) point out in relation to children's knowledge of Binding Principle B that performance in a GJ or comprehension test does not have to be adult-like in order to show adult-like knowledge. That is, to show knowledge of Binding Theory children need only treat grammatical sentences systematically differently from ungrammatical sentences. In Grimshaw and Rosen's own experiment children coped successfully with grammatical sentences but performed only at chance on ungrammatical sentences: that is, although they did not reject them, they clearly recognised them to be unlike the grammatical items. As with second language learners, the reduced lexicon of the child may obscure the full extent of her grammatical awareness.

As we shall see, some of the studies discussed here are not at all exact - for reasons summarised in 3.3 - about their subjects' level of proficiency. Some studies clearly include L2 speakers who could not pass for native speakers. We may then be looking not so much at two clearly demarcated groups, an exceptional elite and the main body of adult learners, as at a cline expressing *degrees* of nativeness, perhaps at more than one level of language

knowledge. Thus the gap between knowledge and performance identified for children by Grimshaw and Rosen (1990) may underlie the striking phenomenon (sometimes termed the Conrad Phenomenon) of L2 learners with far from nativelike proficiency in spoken English producing nativelike intuitions in GJ tests (Patkowski 1980, Snow and Meijers 1977). But it may also be a feature of all adult L2 attainment, underlying for example the less marked differential achievement in phonology and syntax observed for even the most proficient adults, and other uneven attainment, including nativelike grammatical knowledge paired with non-nativelike (albeit advanced) grammatical proficiency in performance.

The studies discussed below all confront Long's challenge in his 1990 review of research into age-related differences: "to determine whether the very best SL [second language] learners actually have native-like competence" (281). Two (Coppieters 1987; Sorace 1993) conclude that non-native competence does not match the native competence; three (Birdsong 1992, Ioup et al 1994, White and Genesee (1996) propose that it can. None however argue that the L2 speakers they have investigated show or are capable of no deviation from native speakers in their intuitions. On fossilization, then, the jury is still out.

3.2   Different Kinds of Competence?

We thus arrive at the key questions not just for ultimate attainment studies, but for SLA research as a whole. Is it possible for *any* non-native to achieve native-like competence? If even the most proficient second language speakers have a different mental representation of the target grammar, what order of differences are we talking about? Recall the 4 different positions on the role of UG in second language acquisition summarised in the Introduction. They can be presented schematically as in Figure 3.1:

UG

	+	-
	A	**B**
L1 +	UG only via L1	no UG - only L1
	C	**D**
-	direct access to UG	no UG - no L1

Figure 3.1 How UG and the L1 interact

D can be considered untenable for adult learners – there is no reason to think that adults at least do not build hypotheses on the basis of L2 input alone. B proposes that adults have no access to UG for L2 learning; and C proposes the opposite - adults have direct access to UG, just as child L1 learners have. A however (access to UG but L1 knowledge as well) admits of three possible outcomes: complete, incomplete and divergent knowledge of the L2:

Hypothesis 1
Native-like speakers can have native-like competence. All learners are in principle capable of success, and lack of success is due to socio-psychological factors such as social distance and lack of motivation which can limit exposure to the L2.

Hypothesis 2
Successful adult learners can only achieve an incomplete version of the native grammar. Positive input is insufficient to ensure restructuring in all cases, and although some parameters are resettable, the end state is an uneven (ie incomplete) competence in non-native speakers.

Hypothesis 3
Native-like performance does not necessarily entail native-like competence. Some parameters are set at intermediate values, or are skewed towards L1 values, without violating UG. As in the Full Transfer/Full Access model (Schwartz and Sprouse 1996), adult access to UG is mediated by learnability considerations (which derive from L1 knowledge), and the resultant L2 grammar, although still possible within UG, diverges from a native grammar.

3.3   Operationalising Ultimate Attainment

A difficulty that has beset all empirical work in this area is that of determining who has and who has not achieved a near-native competence. A related issue, given that we are discussing the ultimate attainment of all L2 learners, not just Selinker's 5% whose achievement is exceptionally successful, is to fix a limit to acquisition, when learners can be said to have reached asymptote.

As we shall see, in most cases stringent objective criteria in subject selection have not been deployed in experimental studies of near-native speakers. This has been mainly due to the absence of any standard test for near-natives: tests for people who can pass for native speakers in most areas of daily life would be redundant, or - perhaps - reveal only stylistic differences in more restricted domains (eg academic writing). When selecting subjects for studies of grammatical knowledge (phonological studies can use rather more rigorous methods), researchers have had to resort to more or less subjective procedures. For the first three studies described below, the experimenters chose their subjects themselves on the basis of informal conversations, or the recommendations of friends. The White and Genesee (1996) study took a more sophisticated approach in which independent judges graded each level of language (phonology, morphology, syntax, choice of vocabulary, fluency and overall "nativeness") against a native-speaker norm. Others (Coppieters 1987, Birdsong 1992) have made use of self-assessment, although L2 speakers do not always reliably assess their own proficiency (Blanche and Merino 1989).

Any procedure, however, has to contend with the avoidance strategies commonly used by highly proficient non-native speakers: if the latter manage to avoid using a particular construction, performance will obviously appear native-like, and differences in TL knowledge will remain hidden, as Coppieters (1987) and Sorace (1993) note. Questions of the significance of a foreign accent (does it preclude a nativelike grammar?) and of domain limitations (eg adult learners are less likely to be familiar with child and school codes) also interact with any assessment of nativeness.

The question of when acquisition in the L2 context ends has received scant direct attention. Hurford (1991:196) describes language as "a finite bounded system which can be known in

its entirety by a speaker". Clearly, however, vocabulary acquisition continues, in both the L1 and the L2, throughout life, as does the learning of different stylistic conventions. Birdsong (1992: 706, fn) argues - "[u]ltimate attainment is therefore not an end point in any simple sense, but an asymptotic curve approaching some theoretical point of absolute completeness." However, acquiring the language system can be detached from accumulating lexis. Some SLA studies have argued that L2 acquisition reaches asymptote after 3 or 5 years immersion in the L2 environment (Johnson and Newport 1989;1991). Studies using production data suggest that even over a prolonged period such as 10-20 years after the setting in of fossilization there is no modification of the interlanguage grammar in the direction of the target L2 grammar (eg Lardiere, 1998). Nevertheless, longitudinal studies of proficient asymptotic learners barely exist, if at all, and ultimate attainment is more a theoretical construct than an empirically verifiable object.

With these caveats in mind regarding the difficulty of assessing near-native proficiency and the absence of adequate empirical evidence for a "final state" of L2 knowledge, it must nevertheless be recognised that ultimate attainment studies have opened up a new route towards the understanding of SLA.

3.4   Ultimate Attainment Studies

Five studies - Coppieters (1987), Birdsong (1992), Sorace (1992), Ioup et al (1994) and White and Genesee (1996) - have investigated the language knowledge of near-native speakers, most but not all within the framework of the Principles and Parameters model. Most research using this model has selected subjects who are in the midst of acquisition (and usually in classrooms), and research, as shown above, has been concerned with the availability or otherwise of UG at different stages. This large body of research has, for rather obvious reasons, concentrated on how the second language is acquired and has not in the main shown so much interest in speakers who have completed their learning, or who have succeeded in achieving - exceptionally for adults - near-native proficiency. These five studies however look at non-natives whose learning may be regarded as complete and who, as far as performance or language use is concerned, are almost indistinguishable from native speakers.

We discuss the five studies referred to above in the light of the three hypotheses given above. Two of the studies (Coppieters 1987 and Birdsong 1992) are not exclusively concerned with parameterized features of core grammar, but with language-specific knowledge as well. However, they accept both the argument that UG is accessed by successful learners at least, and the separate status of UG knowledge. It is therefore feasible to make a comparison of these five studies - on this and on empirical grounds.

By examining SLA through the other end of the telescope - from the end of the process rather than from the beginning - these investigations open up the possibility of:

(1) more fine-grained analysis of the relationship of input to parameter-setting;

(2) new conceptions of bilingual competence.

3.4.1. Coppieters (1987)

Coppieters' study, "Competence Differences Between Native and Near-Native Speakers" - the first in this domain - has the explicit aim of discovering the boundaries of variation in competence within a language community. He asks whether knowing a language, as a native or a non-native speaker, implies sharing a specific range of intuitions about the language ("Does a language impose a grammar on its speakers?" (Coppieters 1987:544)) or if it is possible to appear native-like with a different set of intuitions. In other words, does the interaction of input with "some cognitive learning mechanism"(presumably UG?) produce, finally, a *uniform* competence?

Coppieters interviewed and tested 21 near-native speakers (NNSs) of French, from a variety of language backgrounds - Romance (7), Germanic (7), Oriental (5) and Farsi (2). They were selected for native-like proficiency on the basis of reports by friends and colleagues, and from their performance in the interviews. Although many had studied French at school, none had had intensive exposure to French before the age of 18. They had spent an average of 17 years living in France (range: 5.5-37 years). All were highly educated (17 were professors). 20 native speakers of French (NSs), with more varied socio-economic and educational backgrounds acted as controls.

The subjects were presented with 107 test sentences, 31 of which dealt with 4 "basic grammatical contrasts" in French, including sentences (from the linguistic literature) that

NSs disagree on (though not ambiguous sentences). In fact, these items included both semantic (eg tense/aspect distinctions) and syntactic (eg cliticisation) features. Subjects were asked to give non-normative "individual, spontaneous and intuitive reactions" (551), without evaluating, justifying or correcting the sentences. The other sentences contained 5 UG-type constructions and subjects were required only to give "straightforward well-formedness judgements" (550), without further probing by the interviewer. Subjects had a three-way choice: "correct or good, uncertain or problematic or incorrect or bad" (553:fn).

The quantifiable data gathered in the experiment was used to establish a "native norm", based on the majority opinion of the NSs. Subjects' divergence from this norm on all sentences was then calculated. NSs were found to diverge on 5-16% of the test sentences; NNSs on 23-49%. There was no overlap between the two groups - the most native-like of the NNSs was 7 percentage points away from the most divergent NS. For the NNSs there was most divergence from the norm for the contrast sentences (35.7%-41.5%), while the UG-constructions showed considerably less (19.4%-33.2%). This was *not* true however of the NSs: their pattern of divergence was quite different (2%-11.8% and 5%-14.3% respectively) (554)

The NNSs' interpretations of the sentences elicited in the interviews showed that they had either very different explanations of the meanings of the contrasted items from NSs or that they were unconscious of different meanings. The UG-type constructions produced, as a whole, smaller differences between NSs and NSSs. It is worth pointing out that, in line with other research, on one UG-constrained feature - A-over-A - the two groups agreed more on violations of the constraint than on acceptable items. On the other hand, there is considerable variation between the different L1 groups on UG sentences, notably on A-over-A. Procedural factors (the small number of speakers in each group; the rather small number of exemplars of each feature) may help to explain this (see Birdsong's critique, below).

Coppieters' conclusion is that NNSs have rather different sets of intuitions about the target language to those of NS, and, further, that the most significant differences "involve not so much the formal areas of grammar traditionally covered under the term UG ... as those typically addressed by linguists interested in 'functional' or 'cognitive' aspects of grammar" (565), ie peripheral rather than core aspects.

It is argued that the NNS subjects have most difficulty with the semantic aspects of contrasts in French grammar because they are accustomed to interpreting sentences by referring to context, and so find the meaning of the decontextualised sentences presented in the experiment very hard to determine: "it may be difficult for a NNS (particularly one whose native language does not formally mark the category or distinction in question) to separate contextual from grammatical information"(567) in everyday speech. Furthermore, the meaning contrasts of grammatical "pairs" which are so clear to native speakers may derive from the "contextually rich environment in which they were initially learned". It is precisely these items which are context-dependent and "particularly resistant to predictive analysis" (568). Coppieters points out that these aspects of grammar are hidden, not such as would show up in performance. An explanation in terms of affective factors is unlikely: many subjects had no trace of a foreign accent, so were presumably happy to pass for French.

Coppieters' study attracted a great deal of attention as the first study of its kind, but also some criticism, largely for its methodological flaws. Birdsong (1992), whose evaluation is the most thorough, points out that the contrast between NNS intuitions on +UG and -UG items - the heart of Coppieters' article - is predicted more accurately by the research procedure than by the items themselves. Coppieters himself notes that the greatest divergences appear in "precisely those [sentences] on which the people interviewed had a chance to express qualitative judgements on possible meaning differences" (Coppieters 1987:554)! When the experimenter questions a subject about his or her intuitions, it is only too likely that he or she will pick up clues about the experimenter's intentions, and will think up different contexts, think more deeply about the issue, etc and give different judgements to those given in a straightforward grammaticality acceptability task. The latter is likely to be free of such "experimenter effects". The results of the two different procedures cannot, Birdsong maintains, be validly lumped together to produce a single table of divergence, as they are not comparable.

Coppieters' characterization of the items as UG or non-UG is also challenged by Birdsong. If 'ce' vs. 'il/elle' and 'à'/'de'+infinitive are included amongst the +UG items, in line with Pollock (1983), the claim that NNS are less divergent on UG features fails immediately. In addition, the number of exemplars of each feature is often so small that discussion of percentage differences is misleading.

The subjective and impressionistic procedure used in the study to select highly proficient subjects has also been the target of much criticism. It is argued (Birdsong 1992; White and Genesee 1996) that as stringent a selection procedure as possible is essential in a study of this kind: it is not impossible that avoidance and other communication strategies, combined with proficiency in restricted and familiar contexts can give unwarranted illusions of nativeness. Other problems are that the educational backgrounds of the native speakers and nonnative speakers are not comparable, and that the NNS group includes professors of linguistics, whose intuitions are hardly naive.

Coppieters' results are therefore not entirely convincing, although they cannot be dismissed. The UG/non-UG contrast in the data may indeed be an artefact of the research methodology. Nevertheless, the fact that there was no overlap in judgement between native speakers and highly proficient (at least) L2 speakers is strongly suggestive that the latter do have mental representations of the target grammar that are different in some way, but which still allow them to pass for native. Could this distinction apply to native-like child learners? Coppieters' findings do not allow us to decide (cf Mack 1984, Hyltenstam 1992).

3.4.2. Birdsong (1992)

Birdsong's intention is to replicate Coppieters' experiment under more rigorous conditions. His study also looks at near-native speakers of French. He asks three questions: (a) Are there competence differences between native and native-like speakers? (b) If there are, '[i]n what grammatical domains do the groups appear to diverge?' (Birdsong 1992:707) and (c) Is UG available in adult SLA? Another research question concerns the effect of age of arrival on grammaticality judgements, and this aspect has already been discussed above (Chapter 2).

As well as selecting as respondents people whose first language was English, and matching the educational background of the 20 near-natives (ENS) and 20 natives (FNS) more carefully, he stipulated they should have lived in France for at least 3 years and been first exposed to the target language after puberty (717). More formal assessment was deliberately omitted, firstly because of the inadequacy of testing instruments at this level and secondly because self-assessed mastery of the L2 was to be included as an independent variable in the analysis of results. He also radically revised the tasks. Three tasks were administered. The first required subjects to interpret 20 lexically and structurally ambiguous sentences, using a scalar grading system (5 points) which deliberately focussed attention on subjective

judgements: from "pas du tout acceptable; je ne la dirais pas" (not at all acceptable; I would not say it) to "tout a fait acceptable; je la dirais" (completely acceptable: I would say it). Secondly, 15 sentences exemplifying three meanings of "bien" ("well"; "indeed"; and either "well" or "indeed") were to be judged using a similar scale. The results showed "the performance of ENS is essentially indistinguishable from that of FNS"(738).

The first two tasks were designed to provide convergent evidence for the main task: an acceptability judgement task on 76 written sentences, again using scalar ratings, with a parallel "think-aloud" task, in which subjects' unelicited opinions were taped (the experimenter was not in the same room). The sentences exemplified a mixture of UG and non-UG features. Birdsong included three +UG features not in Coppieters (Adjacency, *that-trace* and Middle Voice), and excluded 5 items from the earlier study. The results showed a significant difference between ENS and FNS on only 17/76 sentences (22%), and, more importantly, no neat pattern of results that conforms to the +UG/-UG distinction" (722).

Table3.1: Comparison of Results from Coppieters (1987) and Birdsong (1992)

LINGUISTIC VARIABLE	+/- UG (Coppieters)	+/- UG (other linguists)	% OF ITEMS WITH SIGNIFICANT DIFFERENCES BETWEEN FNS AND ENS	COPPIETERS' RESULTS
EN-AVANT	+UG	+UG	11	AGREE
ADJACENCY	N/A	+UG	19	N/A
'THAT-TRACE	N/A	+UG	40	N/A
MIDDLE VOICE	N/A	+UG(?)	44	N/A
'DE' + MODIFIER	-UG	+/-UG	25	DISAGREE
'CE'/'IL' ('ELLE')	-UG	+/-UG	0	DISAGREE
PRENOMINAL PAST PARTICIPLE	-UG	-UG	17	DISAGREE

On only one of the items (*en*-Avant) do the two studies agree. Moreover, the way the two groups of subjects in Birdsong's study diverge from native norms is remarkably similar, with considerable overlap between the groups: 15/20 ENSs have deviance scores within the range of FNS scores. It is important to note however that there *is* a statistical difference overall on divergence from native speaker norms between the groups (F= 10.66, *p* 0.002) (724) albeit smaller than Coppieters'.

The think-aloud data are more difficult to interpret. Subjects did not comment on all items, and a few made none. Unsurprisingly given the fact that subjects were not guided as to the kind of comments looked for, it proved difficult to determine what properties of a stimulus sentence were prompting the comment. It seems that FNSs tended more than ENSs to comment on stylistics and comprehensibility. There were clear differences between FNSs and ENSs on several items, (occasionally contradicting results from the numerical data), but the nature of the data (eg too many irrelevant comments) precludes firm conclusions. The results from the first two tasks (ambiguous sentences and interpretations of "bien"), designed to provide triangulating evidence for the main task, do suggest that the two groups did not disagree although there was a high degree of variability in both.

Birdsong concludes that procedure, subject selection and scalar ratings are largely responsible for results so different from Coppieters (1987). He also argues that the difference in modality tends to invalidate comparisons with Johnson and Newport (1989, 1991) whose results were diametrically opposite.

The fact that UG features did not prove a lesser source of divergence than non-UG features conflicts with findings by Zobl (1990) but agrees with Johnson and Newport (1991). Birdsong proposes that although this does not indicate the unavailability of UG, it does "suggest that UG is not accessed fully by all learners in the sample" (740). He also points out that it is unlikely that a bundle of unrelated UG features *would* produce homogeneous results. However, two items which are related under the Empty Category Principle - 'en'Avant and *that*-trace -should produce similar deviance scores, but do not (11% and 40% respectively).

This study clearly produces results different to Coppieters' but, except on the status of UG knowledge in near-native grammars, they partially confirm the findings of the earlier study: NS and NNS differed significantly in the way they diverged from the NS norms. A number of problems of approach and of methodology in Birdsong's research need to be considered and these are discussed below (Chapter 6).

3.4.3. Sorace (1993)

Sorace's study of highly proficient English and French speakers of Italian considers two questions: (a) Are core aspects of the target language represented in the brains of native and non-native speakers in the same way; (b) If not, do the latter exhibit incomplete or divergent

grammars as compared to the TL grammar? Sorace selects a property of the core grammar of Italian which is also a feature of French, though it varies parametrically, and uses it as a basis for comparing the L2 grammars of French and English speakers of Italian.

The study is based on the Unaccusativity Hypothesis and on analyses of restructuring constructions involving the group of Unaccusative verbs in Italian. Unlike English, both French and Italian make use of two auxiliary verbs to form perfect tenses for lexical verbs: "avoir"/"avere" and "être"/"essere". There are differences in auxiliary choice between these two languages: Italian requires 'essere' selection for all unaccusatives while in French only a minority take "être", the rest taking "avoir"; Italian also allows clitic-climbing and restructuring, with obligatory change of auxiliary from "avere" to "essere" in certain contexts. This syntactic paradigm interacts with semantic aspects. Sorace presents a semantically-based Unaccusative Hierarchy with, at the top, verbs expressing change of location or condition (eg "andare" (go), "arrivare" (arrive), "venire" (come)) where "essere" selection is obligatory (core), and towards the bottom verbs (eg "aumentare" (increase), "correre" (run)) for which native speakers find "avere" with a meaning change less unacceptable than with other verbs (periphery). In French, only change of location verbs obligatorily take "être"; change of condition verbs are on the periphery and can take "avoir" or "être"; all other Unaccusatives which require "essere" in Italian take "avoir" in French. French learners of Italian therefore need to take on board a different syntactic and semantic configuration for this class of verbs. English on the other hand has lost all syntactic reflexes of unaccusativity, while maintaining a lexical category of unaccusative verbs. For English learners of Italian then, there is no parameter resetting to be undertaken - a new principle must be acquired.

The subjects were 24 English speakers and 20 French speakers, all judged by the experimenter to be near-native speakers of Italian, and 20 native speakers. They had all acquired Italian after the age of 15, and had 5-15 years of exposure, including instruction, to Italian (average 9 years), not all in Italy. The experiment required them to judge 48 sentences exemplifying three properties of unaccusativity: (a) "essere" selection with five classes of unaccusative verbs along the hierarchy (b) optional auxiliary change and (c) obligatory auxiliary change in clitic climbing. These properties are partially instantiated in French, as described above, and not at all in English. Results showed that near-native speakers do have different intuitions from native speakers, and that, while both NNS language groups were

sensitive to semantic categories in the hierarchy, only the French speakers made consistent judgements on syntactic grounds. However, these judgements diverged from native Italian judgements on verbs lower down the hierarchy, as predicted. English speakers showed - in Sorace's term - *incomplete* knowledge of the target grammar, by giving indeterminate judgements throughout.

All language groups were more likely to accept "avere" as an auxiliary the further down the hierarchy the verb appeared - that is all were sensitive to semantic categories. Where syntactic rules require "essere", as in clitic climbing, only the French judgements corresponded to those of the Italians, while English judgements were indeterminate. Where auxiliary change is optional, Italians accepted both "essere" and "avere", while the French rejected "essere" and the English were again indeterminate:
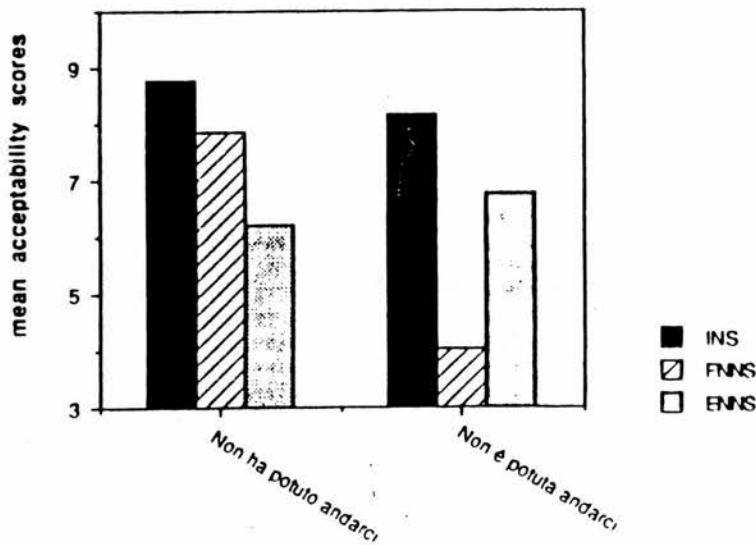
Figure 3.1 Scores of English, French and Italian native speakers on acceptability test of optional auxiliary change (Sorace 1993:41)

The native-like production of Sorace's subjects was not matched by native-like performance in the experiment. Sorace concludes, with Coppieters (1987), that such L2 speakers do not have native-like competence. The semantic distinctions of the Unaccusativity Hierarchy were evident to both the language groups, but they came to different conclusions about the accompanying grammatical rules. Moreover,

the judgements given by the near-native subjects indicate that the use they have made of evidence in the input to modify their interlanguage grammar has been selective, despite the ample availability of positive evidence for all the grammatical aspects investigated. (43)

Fossilization thus seems to have taken place in the *presence* of readily available, quite salient and relevant input. Nevertheless, the differential performance of the two language groups does suggests some L1 effect - that some learners are "in a more favourable position to *notice* the relevance of the L2 input and to incorporate it in their developing grammar"(43). Any L1 effect must be indirect, otherwise the English-speakers would overgeneralize "avere" and the French-speakers reject "essere" for clitic-climbing. Surface transfer is not an issue - rather how unaccusativity is represented in English and French L1 grammars.

Sorace's study, although very different in approach, broadly agrees with Coppieters (1987) and Birdsong (1992), but makes a much stronger case for divergent competence than do either of the earlier accounts, although in only one aspect of grammar. She also measures more precisely the differences between native and non-native grammars. This shows that in the case of unaccusativity, access to the input, at least for adults, is being blocked by prior L1 knowledge - in the case of French speakers, by a different parameter setting; for English speakers, by the absence of the parameter in their L1.

Since a rigorous methodology is essential in this area of SLA, perhaps more than in than in others (because the hypotheses depend on the native-like proficiency of the subjects), it is worth noting that very little information is offered on the subjects. We are not told how well they were matched for age, educational or socio-economic background or even residence in Italy: the single criterion was native-like fluency and accuracy, as demonstrated in conversation with the experimenter.

3.4.4   Ioup, Boustagui, El Tigi and Moselle (1994)

Ioup, Boustagui, El Tigi and Moselle (1994) investigated the performance and competence of an uninstructed non-native speaker of Egyptian Arabic, Julie. Julie had lived in Egypt for 26 years at the time of the study, having arrived at the age of 21. Her spoken production was judged at native speaker level by most members of a panel of linguistically experienced judges, and a grammaticality test and interpretation test revealed her intuitions on a range of UG and language-specific features to be largely native-like (ie close to the intuitions of 11

native speakers).

Causes of this clear breach of maturational constraints have been discussed above. Here we are concerned with how this case enters into the debate on competence in ultimate attainment. Although Julie - and Laura, an instructed second language speaker of Arabic who was also a university teacher of standard Arabic - differed both in judgements and in production (on the translation task) from most of the native speakers in just a few instances (5/37), it is instructive to examine these cases. Firstly, Julie differed where native speakers showed variability in their own intuitions (ie less than 80% concurrence), for example in questions where Julie judged ungrammatical sentences to be grammatical. For UG constructions, Julie's judgements disagreed with the majority on two of the items with the *lowest* concurrence. In these cases both Julie and the native speaker controls may have been responding to flaws in the test materials, some of which (the UG items) were "linguists' sentences", or to the aural modality (necessary because Julie could not read or write in Arabic). Alternatively, there may be optionality in the grammar of Egyptian - the fact that in most of the test items which evoked less than 80% native speaker concurrence Julie sided with the majority and that these items were randomly distributed makes this more plausible. The authors confirm that for question word order (a language particular construction), a majority disagreed with linguists' intuitions, and, interestingly, none were at all undecided about their intuitions.

Secondly, where native speakers allowed more scrambling in questions, Julie was more restrictive. She consistently rejected variable word order in questions, and held to the unmarked (wh in-situ) order. In the third test, on the interpretation of anaphora, Julie consistently adopted one interpretation, disregarding the rule for adjacent and remote reference in relative clauses. The authors suggest that processing difficulty was responsible, as the items were "deliberately complex" (90) but one could equally well point to Julie's lack of formal instruction and literacy in her L2. It must be again emphasized however that instances of disagreement with native speaker norms in the anaphora test were few and that apart from this construction Laura, with a very different learning background, gave similar responses.

Following Sorace's proposal, we can categorise Julie's grammar of Egyptian Arabic as divergent, not incomplete, with respect to the native speaker grammar, as there is no clear

evidence of indeterminacy in her intuitions[1]. However, we cannot conclude that UG constructions were favoured as compared to language particular constructions, as Birdsong also found. It is also unclear what the influence of Julie's L1 was.

3.4.5. White and Genesee (1996)

The most recently published study in this field to be discussed here approaches the question of ultimate attainment from the perspective of the critical period hypothesis debate. White and Genesee are interested not only in whether or not L2 speakers can achieve native-like competence, but in whether or not this depends on the age when intensive exposure begins. Their study examines L2 speakers of English living in Canada.

They contend that all previous experiments can be criticised on the grounds that they use only subjective criteria for respondent selection. Consequently, these studies

> indicate quite clearly that there are age-related effects on second language proficiency. Most of them leave unanswered the question whether late second language learners can ever attain linguistic competence that is indistinguishable from monolingual native-speakers. (235)

They propose therefore to apply much more stringent criteria. They also aim to test for the availability of UG in SLA: if available, there is no reason to resist the hypothesis that NS and NNS competences are similar. Two tests were administered to investigate subject's knowledge of Subjacency and ECP violations of wh-movement constructions. These were a timed, visually presented grammaticality judgement (GJ) test, and a question formation test (written). Another difference from previous studies was that their choice of grammatical test items allowed them to establish whether their subjects had acquired long wh-movement, where J&N (1991) included only a small number of such items (on which their subjects were notably inaccurate).

89 adult L2 speakers were selected, both immigrants and Canadians by birth, with a variety of first languages (58 French speakers, 8 Romance speakers, 8 Germanic speakers and 15 speakers of other languages). First exposure to English varied from babyhood to after 16 years of age (age of arrival groups were: 0-7; 8-11, 12-15; 16+). Their proficiency was

assessed from recordings by three independent judges who scored for pronunciation, morphology, syntax, fluency, vocabulary and overall nativeness. Rather than allocating subjects to points on a scale, the judges marked an 9cm long straight-line rating scale, labelled native at one end and non-native at the other. On the basis of these ratings, the respondents whose overall score fell in the right-most half-centimeter of the line were judged to be near-native, with those rated below classed as non-natives. A control group of 19 native speakers (not from Quebec) also took the tests. All groups were matched for age (NNS: age-range - 16-66 and average age - 29; NS: age-range - 15-73 and average age - 28).

The results of the GJ test were divided into responses to grammatical sentences and responses to ungrammatical sentences as the two types tested for two kinds of knowledge (recognition of grammaticality, and long-distance wh-movement and knowledge of restrictions on wh-movement respectively). No ambiguous or controversial sentences were included. Significant differences between the non-native group and the other two groups were found in both cases, but none between native speakers and near-native speakers. On the ungrammatical sentences, most of the difference between the non-native speakers and the others was due to a single feature, *that*-trace violations, already noticed as producing variable judgements among native speakers. When analysis according to age was carried out, it was found that "within the two groupings, each age subgroup was as accurate on UG violations as every other age subgroup" (251). There was also no significant difference between natives and near-natives first exposed to English after 16, although this was not true of the non-natives. White and Genesee conclude that "L2 learners can achieve native-like competence with respect to constraints of UG, provided care is taken to insure native-like proficiency" (251). Proficiency also has an effect on variance: non-native mean scores on ungrammatical items exhibit nearly twice the variance of native and near-native mean scores but there is no parallel contrast between older and younger learners. Recall that the main feature of the adult learners in J&N (1989;1991) was the very wide range of scores, compared to the more homogeneous pattern found for child learners.

Reaction times on the GJ test were also calculated to ascertain whether the L2 speakers took longer to process the test sentences. Results again showed no significant differences between native and near-native speakers, but significant differences with the non-native speakers.

White and Genesee conclude that their results suggest that "native-like attainment is possible,

regardless of age of initial significant exposure to the L2" (258). Even the non-native group (which was only very slightly less proficient than the near-native group) performed more accurately than in comparable studies. They attribute their very different results simply to the way they selected their respondents, and to the tasks set - and this would also account for the absence of age effects. They acknowledge however that other areas of grammar might not show the same results.

White and Genesee's results are impressively different from those of the other studies discussed here, and would appear to tilt the balance in favour of identical competence, at least for Subjacency. However there *are* statistically significant differences between the non-native group and the other two on most of the GJ items and on the reaction times. Given that they were judged to be only very marginally less proficient than the near-natives, but that their test performance was more than marginally different, White and Genesee may not have completely proved their point.

In view of their claim to more stringency in respondent selection it is worth pointing out that their criteria are no less subjective than those of other researchers, since they rely on impressionistic judgements of stretches of conversational speech. The communication strategies mentioned earlier (eg avoidance, paraphrase, simplification) could equally apply here.

3.5   UG Availability and Native-like Competence

We return to the three hypotheses formulated above. Three of the studies appear to support Hypotheses 2 and 3: adult learners can achieve divergent or incomplete grammars of the target language. Thus Coppieters, Birdsong and Sorace find a measurable difference, albeit small, between the competence of native and very proficient non-native speakers of a language. Ioup et al found almost no difference but their subject, Julie, may belong to a rather different population - that of exceptional or talented learners (other cognitive and psychological features point this way[2]) White and Genesee's findings are radically different, in that they have uncovered a large number of L2 speakers whose intuitions exactly match

those of native speakers. In this case Hypothesis 1 seems justified. However, as the authors admit, only one feature of English was tested in this study, and for most of the subjects there was no parametric variation between their L1 and the L2 settings. It is therefore difficult to claim that White and Genesee's subjects have full access to UG, rather than access mediated by L1 knowledge.

We propose that some form of fossilization may characterise *all* second language acquisition. However, we cannot conclude that native-like competence is a theoretical impossibility. In addition, the evidence adduced so far can neither confirm nor reject the possibility that there is a difference in the competences of near-native speakers which can be attributed to the age of first exposure.

Only Sorace and White and Genesee investigated properties of UG, while Coppieters', Birdsong's and Ioup et al's approaches were more eclectic. Given this, it is still possible to draw some broad conclusions about the status of UG in the competence of near-native speakers. All the studies conclude that such speakers recognise UG constraints. Whether UG grammar is privileged - as Coppieters argued - so that UG features are more easily or more perfectly acquired than peripheral or data-driven language-specific knowledge seems unproven (given the rather random nature of Birdsong's and Coppieters' selection of items). Sorace's results show that even at near-native levels of proficiency, UG-constrained competence does not necessarily mean target-like competence.

The studies described in this chapter seem to bring us no closer to a confirmation of any of the three hypotheses we began with. White and Genesee's work would seem to confirm Hypothesis 1 - that adult learners can achieve native-like competence although they rarely do so. The investigations by Sorace, Ioup et al and Birdsong, together with Coppieters, tend to support Hypotheses 2 and 3: Sorace's English learners of Italian displayed incomplete knowledge and her French learners divergent knowledge; Birdsong's and Ioup et al's subjects seemed to diverge. There are, then, two possibilities as regards the nature of the grammatical competence of highly proficient second language learners: it could be that UG is variably or hierarchically available: some parameters are harder to set than others, depending on the relationship between the L1 and the L2. In consequence, it is the L1, or rather learnability, that determines how complete and how native-like a representation of the L2 grammar is achieved. However, bilingual competence may be necessarily different from

monolingual competence (Sorace, personal communication): that is, acquiring native-like proficiency in a second language restructures language knowledge as a whole, including knowledge of the first language (see also Cook 1990:594)[3].

3.6   Age Effects and Ultimate Attainment

Only two of the studies, Birdsong's and White and Genesee's, included age of arrival as an independent variable, and their findings were rather different. Both found that non-natives' competence can match that of native speakers in certain features. Birdsong found that this was age-dependent: the older the learner was on arrival in France, even after adolescence, the less likely she was to give NS-type judgements. There was thus a linear decline in L2 knowledge. This result contrasts strongly with Johnson and Newport (1989,1991 - see above), who found no age of arrival effect after puberty and considerable variability amongst older arrivers. White and Genesee found no age effect, either before or after puberty, which they attribute largely to rigorous subject selection - that is, they *selected* talented, or at least exceptional, learners. However, they tested only one grammatical feature. It may be, as Ioup et al (1994) maintain, that exceptional language learners can overcome the influence of a critical period. If so, there should be no age effect after adolescence for such learners.

However, the research hypothesis to be advanced here is that older near-native adult arrivals will show evidence that their Interlanguage grammar diverges more from the grammar of native speakers than does that of younger adults, but that it is still constrained by UG - it is still a "possible" grammar. Hence the Critical Period Hypothesis is argued not to hold for such learners. This hypothesis is based firstly on the studies described here which maintain that adult learners develop Interlanguage grammars which comply with UG constraints, though not always with the target grammar, and are hence derived at least partially from a continuing ability to access UG. Secondly, the studies reviewed in 2.4 above suggest that older people's ability to make use of their grammatical knowledge in production or comprehension is increasingly undermined by memory and attentional deficits. For the L1, this may start to happen as early as the fourth decade. It is predicted that the strain placed on cognitive resources when acquiring a second language may, for the older learner, interact

unfavourably with even a slight reduction in working memory to limit the extent of possible grammatical knowledge in the L2. It is thus ageing, rather than maturation, that inhibits the language acquisition of these highly proficient learners.

*

---

[1] A problem with Ioup et al's study is the high proportion of indeterminate NS judgements in the grammaticality judgement test: out of 37 items, only 17 achieved over 85% agreement from the 11 NS informants, while 12 got 70-84% and 8 55-69%. On Subjacency, two Arabic linguists conflictd as well, there seemed to be no satisfactory explanation for this, and the patterns of indeterminate responses, by both the NSs and by Julie seemed quite random.

[2] The Geschwind cluster consists of apparently unrelated factors which tend to appear in individuals with exceptional talents, including language learning talent. They include extreme fairness, homosexuality, schizophrenia, twinning, non-righthandedness and allergies. (Schneiderman and Desmarais 1988)

[3] A third proposal comes from Tsimpli and Smith (1991) who argue that evidence of partial parameter-setting (from their subject, Christopher, a polyglot 'idiot savant') supports the position that only functional categories are parameterised, and that these cannot be reset after puberty, while the principles of Universal Grammar escape maturational limits Thus the kind of grammar an adult learner ends up with must be a "possible", UG-constrained grammar, but it will also reflect the inductive learning strategies adults try to employ in the absence of the option of parameter resetting.

<div align="center">

CHAPTER 4

WH-MOVEMENT: KNOWLEDGE, ACQUISITION AND PROCESSING

</div>

4.1   **Introduction**

4.2   **Constraints on WH-Movement**

4.3   **Barriers**

4.4   **The Subjacency Parameter**

4.5   **The Empty Category Principle**

4.6   **Child Acquisition of Movement Constraints**

4.7   **Second Language Acquisition: Resetting Parameters**

4.8   **Processing Wh-Movement**

4.9   **Lexically-based Accounts of Wh-Movement Constraints**

4.1   Introduction

We have taken thus far a UG-based approach to the questions of grammatical knowledge and acquisition, premised on the capacity of the speaker to utilise input beyond the surface level and thus to achieve a knowledge of a grammar that is more than the sum total of input. In this chapter, we describe two important properties of UG, the Subjacency Condition and the Empty Category Principle, which both constrain wh-movement. We describe how they are acquired in first and second languages and to what extent they are acquired in the latter, and how knowledge of these constraints interacts with processing in real-time comprehension.

In the Introduction, Universal Grammar was briefly defined as a set of principles and parameters which limit the possible sentences of a language and which constitute the innate knowledge of language shared by all human beings. Principles are common to all languages. Parameterised constraints have different "settings" or values for different languages. In Italian, for example the Null-subject parameter is set so that either of the following sentences is possible[1]:

(1a)   Cerco   la scuola
     [I] look for the school
(1b)   Io   cerco   la scuola
    I   look for the school

One way of testing whether learners have made use of Universal Grammar, then, is to discover whether they have "reset" a parameter from their L1 to the L2, where the two languages instantiate different settings: that is, do they accept sentences with the L2 setting and reject sentences with the L1 setting? In the case of an Italian learner of English, evidence that the Null-subject parameter had been reset would be rejection of (1a) in English. There are of course other possibilities. Learners may observe neither the L1 nor the L2 setting, but a third alternative which is still a possible setting for UG. Thus, a naturalistic English learner of Italian might overlearn (1a) and never accept (1b), perceiving "io" as noise.

4.2   Constraints on Wh-movement

Not all languages instantiate wh-movement. Chinese, Indonesian and Hindi, for example all instantiate wh-*in situ*. Elsewhere, extraction possibilities vary crosslinguistically. One constraint on extraction, which varies parametrically between languages is the Subjacency Condition. In what follows, we make use of the accounts of Subjacency and the ECP in the Government and Binding and the Barriers frameworks. These accounts provide the most detailed analyses, particularly of the parametric variations of the Subjacency Condition to be found in different languages. SLA studies using Subjacency and the ECP have also been based on these earlier exegeses. The most recent development of the theory, the Minimalist Program, proposes that "Subjacency violations fail the economy condition that requires chain links to be minimal" (Chomsky 1995: 92), but do not constrain movement at LF. Stronger violations, which are marked at LF, are now covered by the ECP[2]

Focus on the Subjacency Condition has been preferred in grammaticality judgement tests in adult SLA studies of highly proficient L2 speakers because it is not directly available from input, and is unlikely to be explicitly taught. Hence, if positive evidence alone can suffice to induce a switch to the L2 setting, learners with access to UG should be able to obey this condition. It varies parametrically between Italian and English so that if Italian learners are able to reject Subjacency violations in English in the same way as native speakers, considerable support is lent to the position that learners' competence can match that of native speakers and is derived from innate mechanisms, rather than deductive processes. It has been selected for the present study for these reasons. Additionally, as a late acquired feature of first language acquisition, and one that in several SLA experiments is only obeyed by advanced learners, it is especially suited to testing the intuitions of highly proficient non-native speakers. Finally, this feature has been used in a number of the experiments (Bley-Vroman, Felix and Ioup 1988, Schachter 1989, Johnson and Newport 1991, Juffs and Harrington 1995, White and Genesee 1996) with which results from the present investigation can be usefully compared.

In the original account of the Subjacency Condition, Chomsky (1973) describes it as a

condition on movement which therefore applies to all languages which allow movement. In English and Italian, elements such as question words, relative pronouns and noun phrases can be moved from their original D-structure site (the level of syntax which encodes the thematic relations of the sentence) to another site. When such elements are moved, they are said to leave a trace in their original site. Movement can be short:

(2a) What$_i$ did you say t$_i$?

or long:

(2b) What$_i$ did you say [that the woman [who served us] should do t$_i$]?

(2c) How$_i$ did they decide [they would mend the car t$_i$]?

(2d) When$_i$ did he tell you t$_i$ [where$_j$ he was going on holiday t$_j$]?

In (2a) the matrix object is extracted; in (2b) the object is extracted over a relative clause from a verb complement clause; in (2c) there is long distance movement of an adjunct from the verb complement clause; in (2d), adjuncts are moved within the matrix clause and within the embedded clause. The Subjacency Condition blocks certain types of movement. Hence construing *when* with the embedded clause in (2d) results in ungrammaticality. Embedded indirect questions (3a), complex noun phrases (3b), adjuncts (3c) and subjects (3d) are islands for extraction by the Subjacency Condition and movement out of them is barred (partial bracketing indicates islands from which extraction is barred and subscripts represent co-indexing):

(3a) *Who$_i$ did you wonder [when$_j$ they saw t$_i$ t$_j$]?

(3b) *What$_i$ do you know the story [that they told Mary about t$_i$]?

(3c) * What$_i$ does she like to shop [before doing t$_i$]?

(3d) *Which book$_i$ did she say that [the author of t$_i$] was coming to speak?

The account of the Subjacency Condition developed in the theory of Government and Binding (Chomsky 1981) proposes that extraction of wh-elements, such as the initial *who* and *what* in (3a-d), is restricted by certain bounding nodes or maximal projections. In English these are NP and IP: a wh-element can only cross one of these before violating

Subjacency and producing an ungrammatical sentence (only the relevant structure is shown):

(4a) *Who$_i$ did [$_{IP1}$ you wonder [when$_j$ [$_{IP2}$ they saw t$_i$ t$_j$]]]?

(4b) *What$_i$ do [$_{IP1}$ you know [$_{NP}$ the story that [$_{IP2}$ they told Mary t about t$_i$]]]?

(4c) *What$_i$ does [$_{IP1}$ she like to shop before [$_{IP2}$ doing t$_i$]]?

(4d) *Which book$_i$ did [$_{IP1}$ she say [that [$_{IP2}$ the author of t$_I$ was coming to speak]]]?

Who in (4a) is moved from a wh-island - an embedded interrogative clause - and is extracted from the direct object position of the lower verb. *What* in (4b) is extracted from a complex noun phrase, from the indirect object PP. In (4c) *what* is moved from the infinitival adjunct clause, and is the direct object of *doing*. In (4d) *which book* is extracted from the subject of the embedded verb complement clause. All four moved wh-elements are thus arguments (eg *see* is a two-place verb with one external (subject) and one internal (object) argument) and are theta-marked by the verb or the preposition[3]. Their traces are properly governed (see section 4.4 below), so the cause of the ungrammaticality lies in a breach of a movement constraint. Consider (5a), which is grammatical, although - apparently - two IP bounding nodes are crossed:

(5a) Who did [$_{IP1}$ you believe that [$_{IP2}$ they had seen t]?

Here, "who" undergoes successive cyclic movement, as it is possible for the extracted element to leave an intermediate trace (t') in a "landing site" at each revolution, or stage of the cycle:

(5b) $_{CP1}$Who did [$_{IP1}$ you believe [$_{CP2}$ t' that [$_{IP2}$ they had seen t]]]

Stage 2	Stage 1

In Stage 1 *who* crosses one bounding node, IP2 to the Specifier position of CP2, [Spec, CP2], its intermediate landing site; in the second stage, again only one bounding node, IP1 is crossed, and the result is acceptable. In (4a) however, the intermediate trace of *who* cannot occupy the lower [Spec,CP] because the cyclicity condition requires that movement be

successive, and lower cycle movements must be completed before higher cycle movements. Thus *when*, which is restricted to the lower cycle, must be moved before *who* - but by then the lower [Spec,CP] is already occupied. So cyclical movement is ruled out, and a Subjacency violation ensues:

(6) * $_{CP1}$ Who$_i$ did [$_{IP1}$ you wonder [$_{CP2}$ t'$_i$ when$_j$ [$_{IP2}$ they saw t$_i$ t$_j$]]]?

       _____   _____

           Stage 2   Stage 1

(4b), repeated as (7), is an example of extraction from a complex noun phrase, which violates the Complex Noun Phrase Constraint (Ross 1967). Once again applying cyclical movement, Stage 1 is legitimate - only one IP is crossed, but Stage 2 involves crossing NP and IP, and is thus inadmissible:

(7) *$_{CP1}$What do [$_{IP1}$ you know [$_{NP}$ the story [$_{CP2}$ t' that [$_{IP2}$ they told Mary about t]]]]?

       _____   _____

           Stage 2   Stage 1

The Complex Noun Phrase Constraint thus operates exactly like Subjacency, and (7) displays a Subjacency violation.

In (4c), repeated as (8) *what* is extracted from the adjunct phrase, and has to cross two bounding nodes, IP1 and IP2, although IP2 lacks TENSE. In this case too, cyclical movement is ruled out as the landing site for the intermediate trace is filled by *before*.

(8) *$_{CP1}$ What does [$_{IP1}$she like to shop [$_{CP2}$ t' before [$_{IP2}$ doing t]]]?

       _____   _____

           Stage 2   Stage 1

Finally, in (4d), repeated as (9), *which book* is extracted from the subject of the lower clause, and must also cross two bounding nodes:

(9) *$_{CP1}$ Which book did [$_{IP1}$she say [$_{CP2}$ t' that[$_{IP2}$ the author of t was coming to speak]]]?

Stage 2	Stage 1

4.3   Barriers

According to the "Barriers" formulation (Chomsky 1986a) the notion of "barrier" replaces "bounding node". One result of this reformulation is that weak Subjacency effects can be better explained by the theory than previously. As we saw, the earlier theory identified the bounding nodes for Subjacency in English as IP and NP. A barrier is a maximal projection (like IP or NP), but only those maximal projections which are not L-marked can be blocking categories (BC) and prevent movement. L-marking refers to the relation between a lexical item such as a verb and the item's complement which is governed or theta-marked by it. The exception is IP which, although not L-marked, is not a barrier for movement by itself, but only "by inheritance" from another maximal projection such as CP or VP. To this is added the notion of VP-adjunction, which allows a moved constituent to cross only part of a barrier, thus avoiding too powerful an application of the Condition. The Barriers account on its own would predict the ungrammaticality of (10) because VP' is a barrier to movement from t' to the specifier of CP (Chomsky 1986a: 29-30):

(10) How [$_{IP'}$does John [$_{VP'}$ t" [$_{VP}$think [$_{CP}$ t'[$_{IP}$ you [$_{VP}$fixed the car]t]]]]

                              #

Chomsky (1986a) points out that adjunction to VP can account for the acceptability of (10) (#=barrier for Subjacency): adjoining t" to part of VP' saves the construction.

Subjacency can now be defined in terms of numbers of barriers, and the more that are crossed the worse the grammaticality, and acceptability, of the sentence. As Haegeman (1994:563-4) points out, the difference in grammaticality between (11a) and (11b) can be explained more satisfactorily by the newer formulation:

(11a) *[$_{CP1}$ Whom do [$_{IP1}$ you[$_{VP1}$ t" [$_{VP1}$ know [$_{NP}$ the date [$_{CP2}$ when [$_{IP2}$ Mary [$_{VP2}$ t'

                                                          #   #

[$_{VP2}$ invited t]]]]]]]]]?

(11b) ?[$_{CP1}$ Which man do [$_{IP1}$ you [$_{VP1}$ t" [$_{VP1}$ wonder [$_{CP2}$ when [$_{IP2}$ PRO to [$_{VP2}$ t'

#

[$_{VP2}$ meet t]]]]]]]]]?

According to the earlier formulation, *whom* in (11a) crosses three bounding nodes (IP1, NP and IP2), and in (11b) two bounding nodes are crossed (IP1, 1P2). Both therefore are Subjacency violations but GB theory does not provide a metric for calculating the difference in acceptability between them. Under "Barriers" however, NP and IP2 are barriers in the former sentence and result in a clear Subjacency violation, but in the latter only one barrier is crossed (CP2, by inheritance from IP1), so that a weaker Subjacency effect is produced[4]

4.4 The Subjacency Parameter

Subjacency is subject to parametric variation. In Russian, for example, the relevant bounding nodes which constrain movement are NP, IP and CP, in Italian, NP and CP. Thus, while neither English nor Russian allows the following sentence, because two IP nodes are crossed, in Italian a similar sentence is grammatical as only one CP node is crossed:

(12a) * George's new idea [$_{CP1}$ which$_i$ [$_{IP1}$ I can imagine [$_{CP2}$ what$_j$ [$_{IP2}$ you think of t$_i$ t$_j$]]]] will soon be widely publicised

(12b) La nuova idea di Giorgio, [$_{CP1}$di cui$_i$ [$_{IP1}$ immagino [$_{CP2}$ che cosa$_j$ [$_{IP2}$ pensi t$_i$ t$_{j,}$]]]], diverrà presto di pubblico dominio

Italian does not so readily permit extraction from embedded questions, because of a rule against non-echo questions with more than one wh-element (Rizzi 1982:51):

(13a) *Chi ti domandi chi ha incontrato?

(13b) *Who do you think who met?

(13c) ?? Chi non   sai   che cosa ha fatto?
       Who do you not know   what   t   did?

(13d) * Who do you know what did?

Italian and English do however behave in the same way in the case of complex NPs:

(14a) *Questo incarico, che non   sapevo la novità che avrebbero   affidato t a te....
     This   responsibility which I did not know the news that they would have entrusted to you

(14b) *This responsibility that I did not hear the news that they would have trusted you with...

The Subjacency Parameter also interacts with the Null Subject Parameter to produce a variety of other contrasts between English and Italian, as Rizzi (1982) has shown. We summarise below the most significant contrasts between English and Italian.

In English, a Subjacency violation can be avoided by a (substandard) resumptive pronoun. This pronoun is base-generated so no movement is involved:

(15a) ? Mary's new husband, who I can guess what he's saying about us ....

In Italian, the Null Subject allows a similarly substandard but phonetically unrealised resumptive pronoun in the embedded subject position:

(15b) ? Questo incarico, che non so   proprio chi possa avere indovinato a chi è stato
       affidato
    This responsibility,   that I do not know exactly   who may   have   guessed   to whom (it) was
    entrusted

Resumptive null subject pronouns can also "save" potential violations of the Complex NP Constraint in much the same way, when the embedded subject is wh-extracted:

(15c) ? Anna, che non abbiamo sentito la voce   che   lasciava   il marito ...

    Anna   who we had not   heard   the rumour that (she) was leaving her husband

Secondly wh-extraction from infinitival verb complements within wh-islands is legitimate in Italian (Rizzi 1982:65) because the most deeply embedded phrase is not CP (as in (16a) below[5]). The comparable English sentence (16c) is less acceptable, though still an improvement on a tensed lowest IP (16d):

[from Rizzi 82:66]

(16a) ?[6] Mario [$_{CP}$ che [$_{IP}$ non immagino [$_{CP}$ perchè [$_{IP}$ tu abbia deciso [$_{?IP}$ di non

      incontrare]]]]]è

    Mario   that   I don't imagine   why   you should have decided   not to meet   is

    una brava persona

    a   fine   person

(16b)* Mario [$_{CP}$che[$_{IP}$ non immagino[$_{CP}$ perchè [$_{IP}$tu abbia deciso   [$_{CP}$che[$_{IP}$ non

    Mario   that   I don't imagine   why   you should have decided   that   you

    incontrerai]]]]]] è una brava persona

    will not meet   is a fine   person

(16c) ?? Mario, [$_{CP}$who[$_{IP}$I don't understand [$_{CP}$why [$_{IP}$you should have decided[$_{IP}$ not to

      meet]]]]] is a fine fellow

(16d) *Mario, [$_{CP}$who[$_{IP}$I don't understand [$_{CP}$why [$_{IP}$you should have decided [$_{CP}$ that [$_{IP}$

      you will not meet]]]]]] is a fine fellow

Epistemic verbs and verbs of saying, which do not allow embedded subjects of infinitival complements to be left in place, do permit wh-extraction of the subjects, in stylistically marked (formal) sentences. Rizzi (1982:78) relates this to Aux-to-COMP, ie Subject-Aux inversion and Case Theory. Aux-to-COMP is only possible in Null-Subject languages:

[Rizzi's examples]

(17a) Quante persone   ritieni   [e essere in grado di pagare il riscatto]?

    How many people do you believe e to be   able   to pay   the ransom

(17b) *Ritieni   queste persone essere in grado di pagare il riscatto?

    Do you believe these people to be   able   to pay   the ransom

This contrasts with English, which allows both the declarative and the interrogative constructions:

(17c) How many people do you believe to be in a position to pay the ransom?

(17d) Do you believe these people to be in a position to pay the ransom?

Finally, certain other verbs in Italian, like "preferire", take infinitival complements with PRO (PRO VP) which are "invisible" barriers to Subjacency, because, Rizzi explains, they take non-branching CPs:

(18a) ? Una casa grande e vecchissima [$_{CP1}$che [$_{IP1}$ci domandiamo [$_{CP2}$chi [$_{IP2}$preferirebbe

    A large and very old   house   that   we wonder   who   would prefer

[PRO comperare]]]]..

PRO to buy

(18b) ? A large and crumbling house which we wonder who would prefer PRO to buy ...

Compare (18a) with (19a) where the lowest CP is branching and so is a bounding node for Subjacency.

(19a) * Una casa grande e vecchissima [$_{CP1}$che [$_{IP1}$ci domandiamo [$_{CP2}$chi [$_{IP2}$ ritieni [poter

    A large old house   which   we wonder   who you believe   to be able

comperare]]]]]

to buy

(19b) ? A very old big house which we wonder who you believe to be able to buy

Rizzi also points that (19a) is unacceptable because Pro-drop cannot apply in infinitival constructions, so there is no possibility of an unrealised resumptive subject pronoun.

The variability in the acceptability of the Italian sentences is not found in the corresponding English examples due, in the main it would seem, to the fact that English uses a "lower" bounding node for Subjacency than Italian (IP, not CP). The "Barriers" reformulation does not of course affect the parametric variation between the two languages, but it does clarify the difference: restricting the parameter to the lowest tensed clause, English adds an extra

barrier, IP, as in (15c) which degrades the acceptability compared to Italian, as in (15a) (Chomsky 1986a:37-39).

To sum up, core violations of Subjacency, such as extractions from adjunct and relative clauses, are common to Italian and English (although not all movement languages disallow the former - see Akan, below), while extractions from wh-islands are weaker violations, and there is parametric variation between the two languages. Wh-island violations can therefore be described as on the periphery of the grammar of wh-movement, and it can be predicted that: (a) such features will be harder to acquire for L2 learners whose L1 lacks wh-movement; and (b) where both the L1 and the L2 instantiate wh-movement but the L2 is more restrictive, L2 speakers will be more likely to prefer the L1 setting. In the case of Italian, this contrast is of the type of (12), repeated here:

(20a) * George's new idea [$_{CP1}$ which$_i$ [$_{IP1}$ I can imagine [$_{CP2}$ what$_j$ [$_{IP2}$ you think of t$_i$ t$_j$]]]] will soon be widely publicised

(20b) La nuova idea di Giorgio, [$_{CP1}$di cui$_i$ [$_{IP1}$ immagino [$_{CP2}$ che cosa$_j$ [$_{IP2}$ pensi t$_i$ t$_{j,}$]]]], diverrà presto di pubblico dominio

4.5   The Empty Category Principle

The Empty Category Principle (ECP) requires all traces left by the moving of NPs or wh-elements to be "properly governed". It thus acts as a constraint on movement, although the ECP refers to the government of traces not to the movement itself. Object traces can be properly governed by the verb - the head of VP, hence *head-government*, through theta-government, or the assignment of a thematic role (goal, patient etc):

(21) What$_i$ do you think we [$_{VP}$ [$_V$ could have t$_i$ for tea]] ?

Subject traces are not theta-governed because they are governed by INFL which as a non-lexical category cannot assign theta roles. Instead they may be *antecedent-governed* through

coindexation with a maximal projection (Haegeman 1994:445-446), as in (22a). However antecedent-government may be blocked, notably by an element in COMP (21b)

(22a) Who$_i$ do you think t$_i$ will come for tea?

(22b) *Who$_i$ do you think that t$_i$ will come for tea?

 Adjunct traces are not head-governed either and antecedent government can be blocked:

(22c) How$_i$ do you think Bill will mend the car t$_i$?

(22d) * How$_i$ do you wonder whether Bill will mend the car t$_i$?

Rizzi (1990) reformulated the ECP as a unified account of why arguments but not adjuncts can be extracted (Rizzi 1990:3). Rizzi's conjunctive formulation of the ECP requires that all non-pronominal empty categories (ie all traces) be both properly head-governed (= formal licensing) and either theta-governed or antecedent-governed (= identification). Argument variables (traces) "are allowed to bear a referential index, which makes it possible for them to be connected long-distance with their operator through a binding relation" (Rizzi 1994:364). An adjunct variable cannot be referentially indexed because it lacks any theta role: the effect of relying only on government relations to connect an adjunct variable to its operator is to make an ECP violation appear stronger in this case.

When ECP violations, which tend to be less interpretable than Subjacency violations, are added to the latter, there is a considerable decrease in acceptability in English:

(23a) *I saw the woman$_i$ [$_{CP1}$ whom$_i$ [$_{IP1}$ you told me [$_{CP2}$ when$_j$ [$_{IP2}$ you

knew t$_i$ t$_j$]]]]

(23b)**I saw the woman$_i$ [$_{CP1}$who$_i$ [$_{IP1}$ you told me [$_{CP2}$ when$_j$ [$_{IP2}$ t$_i$ knew

you t$_j$]]]]

In (23a) the object trace in the lower clause is head-governed by the lower VP, and so only a Subjacency violation follows (two IP nodes are crossed by the extracted object). In (23b) the subject trace cannot be head-governed by the lower IP, and antecedent government by *who* is

blocked - by the maximal projection CP2.

Extraction of objects, which are theta-governed as in (23a), leads only to Subjacency violations when the relevant bounding nodes (or barriers) are crossed. Extraction of subjects (23b) has to be worse than object extraction because they are not directly theta-marked under head government, although they can have referential indexes, and in long wh-movement antecedent government may be blocked, as we have seen. Although extraction of subjects tends to be as unacceptable as extraction of adjuncts, when more embedded clauses are added this symmetry seems to disappear and subject extraction becomes more like object extraction (although this may not be connected to the ECP (Rizzi 1990:81)). In fact the main distinction is between arguments and adjuncts: the least acceptable violation is extraction of an adjunct (=ECP violation) from an adjunct or a wh-island (=Subjacency violation):

(24) ***Where$_i$ did Ann see [before visiting t$_i$]?

There also seems to be a difference in acceptability between locative and temporal adjuncts like "when" and "where" on the one hand, and manner and reason adjuncts ("how" and "why"). This may be due to the fact that events referred to by verbs take place in time and space - verbs in some way select for specific times and places (see Rizzi's account of adverbial PPs, Rizzi 1990:91). Time and place adjuncts are thereby referential (closer to arguments) and extractable, while manner and reason adjuncts are not.

The ECP seems not to vary parametrically between languages. There is however one difference between English and Italian as regards the ECP, identified by Rizzi (1990:72-76). Compare (25a) and (25b):

(25a)* Which train$_i$ do you wonder when$_j$ t$_i$ would arrive there t$_j$?
(25b)? Quale treno$_i$ ti domandi quando$_j$ arriverebbe t$_i$ t$_j$?
      which train you wonder   when   would arrive

Italian - along with other Null Subject languages - allows a postverbal subject position, thanks to the possibility of subject-verb inversion (see above). In postverbal position the

subject, and hence as here the subject trace, can be head-governed by INFL (and theta-governed by V^7). This fulfils the conjunctive formulation of the ECP. The marginal acceptability of (25b) can be attributed to a mild Subjacency effect. Similarly (23b) in Italian could include a subject trace, thus avoiding an ECP violation, though not a Subjacency effect:

(23b)* Ho   visto la donna chi mi hai detto quando ti conosceva t
      I have   seen the woman who you told me   when   knew you

In addition Italian apparently does not require the *that*-trace filter found in English (Rizzi 1982:145-154). (26b) is unacceptable, according to the ECP because the subject trace is not properly governed, while (25a) shows no ungrammaticality:

(26a) Chi credi che verrà e?
(26b) *Who do you believe that t will come?

Rizzi (1982:146-147) again appeals to the Null Subject Parameter which permits rightward movement of the subject and then wh-extraction from postverbal position. Hence *that*-trace effects in Italian are avoided rather than absent.

The status of "that"-trace in English is somewhat in dispute. In acceptability judgement tests it has frequently elicited fairly indeterminate responses (Bley-Vroman 1988, White and Genesee 1996). The fact that it can be deleted at LF, and that the effect is suspended if a sentential adverbial intervenes between "that" and IP leads Culicover (1993) to argue against an ECP account.

4.6   Degrees of Grammaticality and Acceptability

It has already been shown (Section 4.3) that the more barriers that are crossed the greater the ungrammaticality as regards Subjacency, and this holds for English and Italian. It can be

predicted therefore that (27a-e) will elicit judgements of decreasing acceptability:

(27a)? What do they imagine how Jim has invented t t?

(27b)* What have you heard the news that Jim is going to invent t?

(27c) **What do you think they will launch the sales campaign after Jim has completed t?

(28a) **Who do you know about the product that t invented?

(28b) ***How do you imagine Jim will retire after inventing his gadget t?

(27a) involves extraction of the direct object of an embedded wh-clause, or wh-island. One barrier is crossed, at the lower [Spec,CP] position, so a weak Subjacency violation ensues. (28b) shows extraction of the direct object from the complement of a factive NP (see 4.8 below), where again one barrier is crossed, resulting in a weak Subjacency violation, but where the choice of "the news" rather than - say - "a rumour" worsens the sentence. In (27c) and (28a) two barriers are crossed in extractions from an adjunct clause and a relative clause respectively. In addition, the extraction of the subject of the relative clause in (28a) produces an ECP violation. Finally, (28b) involves moving an adjunct out of an adjunct clause, which results in the least acceptable sentence (a strong Subjacency violation together with a strong ECP violation).

It can also be predicted that the more ungrammatical a sentence - ie the more barriers that are crossed - the more unanimous will be the judgements of native speakers. Conversely, sentences like (27a) which exhibit only weak violations are liable to elicit more indeterminate, that is more variable judgements. Judgements of properties that vary parametrically between languages are evidently less robust than those which do not [8], and in this case it is the grammar itself, rather than individual responses to it that produces indeterminacy (Sorace 1990:133).

Individual native speakers may be quite certain of their judgements (see Ioup et al (1994)), but may disagree among themselves. Non-native speakers from either movement or non-movement languages will also give more indeterminate judgements of sentences like (27a) than of sentences lower down the grammaticality hierarchy - in this case, the indeterminacy may be an indication of absence of certainty as to their grammatical status, produced either by developmental factors (eg later acquisition, absence of input) or by L1 influence.

In the experiment undertaken for this thesis Italian speakers of English were presented with sentences which instantiate wh-movement out of NP complements, embedded interrogative clauses, and relative clauses, violating both the Subjacency Condition and the ECP, in order to elicit their knowledge of the grammatical possibilities of movement. They were also shown sentences which violate the Subjacency condition in English by instantiating the Italian setting of the parameter. It was predicted that such sentences, together with wh-extraction from embedded infinitival clauses, would prove more acceptable for Italian speakers of English than for native speakers.

Assuming that the ECP does not vary parametrically, but is an invariant principle of all movement languages, and that "Barriers" permits differentiation between strong (= 2 or more barriers crossed) and weak (=1 barrier crossed) Subjacency violations, we can identify a core and a periphery within English and Italian for constraints on wh-movement, with the former containing stronger constraints and the latter weaker constraints. An "acceptability hierarchy" of wh-movement violations for Italian learners of English, based on such a core and periphery typology would predict that the higher up the hierarchy the more likely it is for such speakers to fail to reject violations. Figure 4.1 illustrates this hierarchy:

CORE	acceptable *What did you say that the woman who served us should do?*
PERIPHERY - ITALIAN SETTING	acceptable *He painted the village which I wonder whether you know?*
PERIPHERY -ITALIAN SETTING	marginally unacceptable *?? The man who I can't understand why you should have decided not to meet is a perfectly nice chap.*
PERIPHERY	fairly unacceptable **What do you know the story that Mary has seen?*
CORE	unacceptable ***What were you sleeping when they saw?*
CORE	unacceptable *** Who do you know the author wrote that book?*

Figure 4.1: An acceptability hierarchy of movement violations for Italian learners of English

4.6 Child Acquisition of Movement Constraints

Children seem to assume initially that English is a non-movement language, like Chinese or Hindi. Research in the late 70s and early 80s into child acquisition of movement rules suggested that children first form relative clauses by conjunction, not movement (Tavakolian 1981). Wh-*in situ*, as in (29) is produced very early (22 months), although alongside moved wh-utterances, which at this stage are probably holistic chunks ("What's that?"):

(29) You know what?

By the age of 3, children are obeying constraints on movement out of temporal adjunct clauses (Goodluck, Foley and Sedivy 1992) and by 4 are starting to recognise in comprehension tests that extraction from complex NPs is restricted by Subjacency (Otsu 1981). Otsu (1981) tested children aged 3 to 10 using a story and picture matching procedure. The children were first told a series of mini-stories without pictures. Each story was then repeated by the experimenter with pictures and the child was asked a question that could be answered either by an answer that obeyed the Subjacency condition on object extraction from relative clauses, or by one that broke it, both being equally plausible:

(30)   *stimulus*: Jane is drawing a monkey with a crayon

         The monkey is drinking milk with a straw

         questions: *What [is Jane drawing [a monkey [that is drinking milk with t]]]

         What [is Jane drawing [a monkey [that is drinking milk]] with t]]

For a child with access to the condition, the only possible answer is "a crayon" Otsu followed this test with two syntax tests, one toy movement and one repetition, to find out if the children had acquired relative clauses. The use of two tests rather than one increased the reliability of the results as the children performed better in the repetition task than in toy movement. It is important to note that Otsu made sure that his child respondents had an adult-like grammar of relative clauses before identifying their responses to the test question as observing Subjacency, or otherwise. That is, children who do not form relatives by movement (although they may form questions this way) do not have wh-islands in their

grammar, and thus do not have a grammar for which Subjacency is relevant.

Goodluck, Foley and Sedivy (1992) also tested very young children (3-4 year-olds) with a similar procedure, on whether they observed the constraint on extraction of objects from temporal adjunct clauses. The children were told a short story, with a picture, followed by a question with two semantically plausible answers: only one, where the question word refers to the matrix object, is syntactically permitted. The wording of the stories was manipulated to avoid performance distortion, and in order to eliminate the effect of processing difficulties, a follow-up question was asked to distinguish between performance and competence errors. Their 1992 study is a replication of Goodluck et al (1989), and the results confirm the findings of the earlier study. They conclude from the two studies that children's knowledge of movement constraints develops earlier than Otsu found. If the difference is not due to experimental design features[9], it may be that "movement as a syntactic operation may develop in a piecemeal manner, at different times in different constructions" (185) and in different children (189).

The acquisition picture is still patchy and in places confused, as discourse, pragmatic and lexical factors are more likely to intervene in comprehension tasks than is the case with adults. Children are also less likely to be able to comment metalinguistically on their performance. Thus Crain and Thornton (1990) argue that children as young as 2 have mastered relative constructions, but that in many studies this knowledge is masked by processing difficulties, probably related to short-term or working memory limitations (see also Saah and Goodluck 1995; Goodluck, Saah and Stojanovic 1995). For example, children are more likely to violate syntactic rules than pragmatic rules, if faced with a choice. Acquisition of wh-movement and its constraints is a more complex enterprise than other features, and seems to proceed by stages each one being a "possible" UG-constrained grammar. As Goodluck and Rochemont (1992) note, more robust constraints like the temporal island condition and relative clauses are acquired earlier, while the complex NP constraint and wh-island vary crosslinguistically, and this variation can emerge in the child's developing grammars. Thus, studies of children's question formation have shown that some produce questions with a medial wh-word in English, which is grammatical in some German dialects (Goodluck and Rochmont 1992:21).

Any account of the route children follow in acquiring knowledge of movement must also consider learnability issues: (how) do constraints such as Subjacency enable children to acquire knowledge of movement despite the deficits and traps for the unwary in the input? Or - to turn the question on its head - what evidence can be adduced to account for the fact that children seem to make selective use of input in developing grammars which, crucially, are restricted enough to allow them to evolve into adult grammars. Wexler and Culicover (1980), amongst others, have produced powerful proposals in favour of principles and parameter as tools for acquisition. The Subset Principle can, it is argued, provide an important key to the puzzle of acquisition without negative evidence. It declares that when one language is a proper subset of another, the acquisition procedure will always guess "the smallest language that is also compatible with the positive evidence so far encountered" As the child hears more utterances, she will alter her initial hypothesis accordingly. The Subset Principle also applies to parameter setting. A study by Roeper and de Villiers (1990) provides evidence that 3-4 year-olds allow long distance movement in wh-questions. They argue that children filter out misleading input such as (29), or at any rate avoid generalising from it, and demonstrate that children allow extraction from adjunct clauses when they are still forbidding adjunct movement out of small clauses - even though small clauses appear in child language long before complementizers. Hypothesizing - somewhat controversially - that small clauses are L-marked and do not contain CP nodes, Roeper and de Villiers suggest that the syntactic movement parameter, being hard-wired, is triggered before the child has the lexical information for the case module. Only this (intrinsic) ordering of parameters avoids the emergence of a grammar which overgeneralizes from input and which violates the subset criterion. De Villiers et al (1990) conclude, from the same experimental data, that "when LD [long-distant] movement arrives, the barrier constraints are present: in particular, the adjunct/argument distinction, the Empty Category Principle and blocking properties of the COMP node" (293) - even though child grammars do not allow successive-cyclic movement (ie are more restrictive than adult grammars).

4.7 Second Language Acquisition: Resetting Parameters

Within the framework of UG, learning the movement rules of a second language involves resetting parameters. According to the Subset Principle described in the last section, L2 learners like child L1 acquirers, can derive the new values from positive input only. This input will also be incomplete and degraded in many ways. If the L1 is a proper subset of the L2, parameter resetting will be possible. On the other hand, if adult learners do not have access to UG, no parameter resetting can take place.

This section reviews empirical work on L2 acquisition of English wh-movement rules by adults, drawing mainly from research conducted within a GB framework.

The value of testing L2 learners on their knowledge of the Subjacency condition and other constraints on movement such as the ECP in the target language has already been referred to in connection with both age of arrival and ultimate attainment studies. Here, we are concerned with L2 learners' treatment of Subjacency violations as diagnostic of their acquisition of English wh-movement at different stages of development.

The first important Subjacency study was published by Bley-Vroman, Felix and Ioup in 1988. A large-scale investigation of upper-intermediate Korean learners of English, it neither confirmed nor refuted the adult access to UG hypothesis or the Subset Principle. Their subjects achieved 75% accuracy on the GJ test (native speakers had 92% accuracy), thus well above chance. According to their TOEFL scores, there was a wide range of ability, but no significant correlations with these scores (or with other factors such as age of arrival or length of residence). The main puzzle of these results is: Why did the Korean subjects not perform better? The authors speculate that processing difficulties and psychological factors may have played a role, but subsequent research rather enthusiastically took up the challenge.

Schachter (1990) and Johnson and Newport (1991) also tested Korean subjects, amongst others. Schachter compared Dutch, Indonesian, Chinese and Korean learners, all described as "highly proficient". She found L1 to be the best single predictor of test performance, with the Dutch group achieving native-speaker like results, Chinese and Indonesians performing

worse and the Koreans giving indeterminate and very variable answers. Schachter's conclusion was that learners were only able to recognise Subjacency violations if the Condition was instantiated in their L1 - adults had no direct access to UG. As has been frequently pointed out in the literature, Schachter's test materials included no examples of grammatical wh-movement, hence there was no evidence that her subjects from languages without movement had acquired such constructions. As we have seen (Part 2 above), Johnson and Newport (1991) reached much the same conclusions about the role of the L1 and UG: even asymptotic adult learners are unable to recognise Subjacency violations with native-like accuracy, although they again performed above chance.

What evidence is there that adult learners can acquire target-like representations of wh-movement constraints, whether by parameter resetting or by setting the Subjacency parameter for the first time? White (1988) tested three groups of instructed French learners of English on Subjacency and ECP violations. Two were adult groups at low intermediate/intermediate and upper intermediate/advanced levels respectively, and the third consisted of adolescent school students (average age 15 years). As learners need to have acquired wh-movement and embedded sentences before they can be aware of constraints, test materials must include both complex sentences and sentences with legitimate extractions. Results from four tests showed the adolescents scoring low on the cloze test for proficiency and performing around chance on both grammatical and ungrammatical items - they had not received enough exposure to allow them to distinguish between them, even though French has wh-movement. For the most part the adults, who were at comparable proficiency levels according to the cloze test, produced comparable results (over 75% accuracy). However, one group was significantly less accurate at rejecting wh-island violations which instantiated the French setting (similar to the Italian). White argues that these adults were "making possible errors, that is, accepting forms which are ungrammatical in English but which are not ruled out in principle by UG" (158). The fact that the other group had better results on this type shows - in White's view - that parameter resetting *can* take place, perhaps when "the right kind" of triggering input is available. The adolescents also performed at chance on ECP violations, while the adults gave responses like those of the native speaker controls, except for *that*-trace sentences, possibly reflecting the weaker status of these violations.

A developmental comparison of Indonesian adult learners provides even stronger evidence that input alone can interact with UG to set this parameter (Martohardjono and Gair 1991). Indonesian has movement only in a subset of topicalized constructions. Gaps are filled by base-generated *pro*, which disallows certain object gaps. The prediction was that Indonesians would both allow Subjacency violations and display a subject/object asymmetry in gap constructions (eg rejecting object relatives but not subject relatives). Results of an elicited imitation task on gaps showed that the intermediate group confirmed the prediction, suggesting that at this stage learners are relying heavily on the L1 grammar. The advanced learners performed close to native levels. However, even the intermediate group had results above chance levels in a paced GJ test of Subjacency violations.

Although these two studies agree, *contra* Schachter (1989) and Johnson and Newport (1991), that (a) learners can access UG principles when exposed to sufficient input, and (b) there is an interaction with the L1 at least in earlier L2 grammars, White (quoted in Martohardjono and Gair) argues against the specific proposal of Martohardjono and Gair that learners can make use of principles such as the ban on object *pro* in Indonesian to reject object extractions. According to the data in Johnson (1988) and Schachter (1989), learners tend to *accept* both grammatical and ungrammatical object extractions.

Martohardjono's 1992 study again sought to find evidence that even when learners fail to achieve the accuracy levels of native speakers - as most adult learners do - their response patterns to relative acceptability hierarchies clearly reflect the action of UG: she compared groups of Italians, Chinese and Indonesians on weak and strong Subjacency violations (according to the "Barriers" account of Chomsky (1986)). All groups including the native speaker controls reacted more positively and accurately to subject extractions than to object extractions, as well as to extractions from adjunct and relative clauses over weaker cases, regardless of L1. In Indonesian the contrast works the other way: the results therefore "strongly suggest that the hypotheses learners form with respect to the L2 are independent to some degree of their knowledge of the L1" (Martohardjono 1992:7).

Uziel (1993) is a replication of this study with subjects whose first languages were Hebrew and Italian (both instantiating movement). Uziel's subjects, scoring overall around 75% as in

earlier studies, were significantly more accurate on strong violations than on weak violations, and on subject extractions compared to object extractions. Nevertheless his conclusion, that this *pattern* of results shows that such learners are able to *fully* access UG, seems unwarranted in view of the similarity between English, Hebrew and Italian with respect to wh-movement. The no-adult- access to UG hypothesis would also predict these results for learners from movement languages. Where the languages differ - on wh-island extractions and on *that*-trace violations - accuracy rates dropped to 60%, not far above chance. Given that the subjects are described as advanced, scoring up to 100% on the placement test[10], the accuracy results Uziel obtains seem rather low.

4.8   Processing Wh-Movement

Studies of how people parse and process sentences involving wh-movement have both enriched and developed the syntactic account. Investigations into neuropsychological aspects of parsing these sentences have produced interesting reinterpretations of the Subjacency Condition as a processing, rather than a syntactic, constraint. Others have shown how parsing strategies are derived from grammatical knowledge. The scope of the present study did not permit an investigation of processing. Nevertheless the evidence from the mass of research into this aspect of language behaviour compels attention and is highly relevant to an understanding of the findings of this and other Subjacency studies.

Most of the discussion that now follows deals with studies of L1 speakers of English: L2 studies of processing wh-movement are fairly sparse to date (Cook 1989, 1990 and Juffs and Harrington 1995 and 1996 seem rare exceptions), although some investigations have included calculations of reaction times (White and Genesee 1996, Murphy 1997). Wh-movement sentences posit the existence of a fronted element (a "filler" - the operator in GB terms), followed by a "gap" (or variable). The fronted element may be subject, direct or indirect object, or adjunct. When the sentence is parsed, the filler cannot, unless it is the subject of the sentence or the clause, be interpreted until the gap is reached. It has been

shown empirically that the processor anticipates a gap in direct object position but not in subject position (see Saah and Goodluck, 1995:402). It has been proposed that it is working, or short-term, memory that allows the parsing of wh-movement sentences by holding the filler until the gap is reached. Thus the greater the distance between filler and gap, the greater the strain on working memory. Kluender and Kutas (1993) measured the electrical activity of the brain (Event-Related Brain Potentials) during the reading of yes/no questions and wh-questions. In this type of study, electrical brain activity is recorded by electrodes placed over different areas of the brain as subjects read sentences one word at a time. They found increased negativity over left anterior regions of the scalp (a left anterior negativity or LAN effect) for the latter compared to yes/no questions, but also a significant difference in the LAN effect for sentences like (31a) compared to (31b) (examples from Kluender and Kutas 1993):

(31a) *Who$_i$ has she forgotten what$_j$ the boss referred$_j$ to$_i$ for further study?
(31b) *What$_i$ have you forgotten who$_j$ had to extricate him from$_i$ when he was younger?

In (31a) both wh-elements are remote from their traces, or gaps; in (31b) the filler in embedded subject position is adjacent to its gap. Sentences like (31b) in Kluender and Kutas' experiment evoked no LAN effect, and grammaticality made no difference to this effect. It is well known that reading speed and comprehension are responsive to individual differences in working memory capacity (King and Just 1991). Kluender and Kutas (1993) suggest that the LAN effect they have detected may be an index of working memory capacity (not perhaps only for linguistic processing) and may not be a direct reflex of filler-gap processing: for example, the absence of a LAN effect for sentences with subject gaps next to their fillers (as in (31b)) indicates that it is a response to the need to hold the filler in memory, not a processing effect. However, research by Neville et al (1991) found a LAN effect based on grammaticality, and this was replicated by McKinnon and Osterhout (1996). According to the latter study, the "remarkably early onset" of the ERP response to movement violations suggests a grammar-based parser, with separate, modularised, processes positing traces.

There is strong psycholinguistic evidence that people tend to posit a gap as soon as possible, and eschew long-distance dependency[11]. Early gap-filling certainly reduces the load on

working memory: in the 1980s, Kimball's Right Association and Fodor's Minimal Attachment Principle both identified the effect of the limited capacity of short-term memory on parsing. Right Association proposes that the parser prefers to attach newly encountered words in a sentence to the lowest possible non-terminal node (which helps to explain the right-branching tendency of English complex sentences). Frazier and Fodor's model (1978) had stated that the maximum number of words that can be parsed at the same time is 6. Schachter and Yip (1990) presented results that showed a processing effect on sentence acceptability. They tested 60 undergraduates (20 Korean speakers, 20 Chinese speakers and 20 native English speakers) and found that for all subjects object extraction was more acceptable than subject extraction. For the NNSs but not for the NSs acceptability varied with the number of embeddings in test items. Schachter and Yip (1990) attribute this latter asymmetry to processing, especially to garden-path effects, rather than to memory limitations. Thus the asymmetry between subject and object extraction is due to parsing difficulty: object extraction allows use of Minimal Attachment, while subject extraction does not.

Anticipating a gap also suggests that people are trying to complete sentences as soon as they feasibly can, that is, as soon as they have identified an interpretable chunk. Of course, feasibility and interpretability depend on the reader's grammatical knowledge. In Kluender and Kutas' (1993) experiment, subjects read sentences on a monitor one word at a time: the LAN effect appeared when the word following the putative gap was processed. In (31b) the word following the embedded subject "who" is a finite verb, indicating canonical word order, and the whole chunk "who had to extricate him" is interpretable; however in (31a) the chunk "what the boss referred to for further study" is less interpretable, unless gaps are projected in object positions. It was in this latter type of construction that LAN effects appeared (Kluender and Kutas 1993: 206)

Kluender and Kutas's working memory hypothesis thus predicts the same processing behaviour as the Completeness Constraint on Binding (Goodluck and Finney 1993), a grammar-based parsing principle which states that "the processor will bind a wh-word to a gap only at positions that are potentially complete sentences" (Saah and Goodluck 1995: 401). Goodluck and Finney (1993) define a "potentially complete sentence" as one in which

all lexical restrictions are complied with, and one which is closed or complete at the level of semantic representation. Measuring reading time (Stowe 1986, Bourdages 1992) shows that readers slow down when they reach direct object positions in embedded questions - ie they expect a gap but don't find it; while this does not happen at subject positions - they don't expect a gap. However, the Completeness Constraint claims that parsing islands, which result when a gap seems to be reached during processing, producing an *in*complete sentence, only mimic syntactic islands - they are not the same as them.

Saah and Goodluck (1995) compared responses to extraction from temporal adjunct clauses by adult speakers of English and speakers of Akan, a Ghanaian language. Extraction from this kind of clause is legitimate in Akan (though strongly prohibited in English), but in two comprehension tests, using materials developed for child acquisition studies in English (Goodluck, Foley and Sedivy 1992), Akan subjects behaved as though Akan followed the English pattern. Only in the third test, a grammaticality judgement test which both employed more adult (and more Ghanaian) test items and allowed some time for reflection, did they accept extractions from temporal islands such as the following:

(32) What did Ama read the Graphic before she wrote?

The authors cite these apparently contradictory results as evidence that syntactic islands are not the same as parsing islands. They define a parsing island as "a position in the sentence that is not considered as a potential location for a *wh-* word during the course of sentence processing" (399). They propose that the Completeness Constraint on Binding accounts for responses that render (32) ungrammatical in rapid parsing - that is, the processor can complete the sentence after "read" by identifying a direct object gap for the wh-word. However, (32) was an item in the third test - an equivalent item from the comprehension tests is (33):

(33) What did Fox eat __(1) before he wrote __(2)?

The gap at (1) is a grammatical site for the trace of "what", and the grammatical answer in this case is "ice-cream" (Goodluck et al's term is the "upstairs" answer). The "downstairs"

answer, ("a letter") is ungrammatical here as it involves extraction from an adjunct, hence a Subjacency violation.

Note that Akan requires a base-generated resumptive pronoun in wh-phrases (overt for animate referents, null for inanimate referents) - a fact that caused difficulties in adapting the test materials. This may, in reality, have been the source of the parsing islands, the effect of which is not apparent when the "upstairs" verb already has a direct object, and there is only one position available for the wh-operator, as in (32). Nonetheless, the fact that the Akan speakers showed such a strong preference for the English setting (83%) rather than giving indeterminate judgements as the grammar would predict does tend to support the parsing island hypothesis. Saah and Goodluck (1995) argue that although the "Barriers" account of the ungrammaticality of extraction from adjunct islands is not entirely convincing (see Lightfoot and Weinberg 1988), native English speakers seem to find such constructions highly unacceptable: "it may be that these clauses are not in fact strong islands in the sense that the constructs of the competence grammar strongly forbid extraction; but the illusion that they are so will come from the fact that such clauses are islands to extraction in processing" (405). However, it may also be that using children's test materials with adults entails different issues of acceptability.

If the processor finds a gap but then encounters a word preventing completion, it is forced to backtrack and reanalyse the gap: hence, if the processor first identifies the gap in (34a) as an object gap, then encounter "likes" it will be forced to return to the gap and read it as a subject gap. Contrast this with (34b), where the first possible gap (according to the grammar) is in the embedded (lower) object position, and this is also the only possible gap, at the end of the sentence.

(34a) Who did Ann say _____ likes her friend?
(34b) Which man did Jane say her friend likes ____ ?

Juffs and Harrington (1995) measuring on-line reading times by Chinese speakers of English on subject and object extractions found even their native speaker controls in fact slowed down more at a subject gap, as in (34a), than at an object gap, as in (34b). They found

accuracy levels on grammatical and ungrammatical object extractions to be similar between native speakers and non-native speakers. However, the significant differences on subject extraction lead them to conclude that the Chinese subjects had acquired long-distance wh-movement, but that the absence of wh-movement in Chinese impeded the *processing* of subject gaps. The greater the number of reanalyses required, the lower the accuracy.

According to Pritchett's Theta Reanalysis Constraint, "as each word comes through the parser, each local string is maximally licensed, that is, the parser attempts to form as complete an interpretation as possible, with all principles (eg Theta Attachment, Case, Binding) satisfied as soon as possible" (Pritchett 1992: 491). Pritchett's proposal argues for the identity of the parser and the grammar on the assumption that the parser is head-driven. Hence, it is knowledge of the grammar that requires that a wh-operator head a chain, and must therefore assign case and a theta-role. Under this approach, parsing islands can better account for some of the facts of Subjacency violations, notably that Subjacency constrains movement only at S level, not at LF (where wh-*in situ* is permitted). Pritchett argues that "island effects are to be attributed to the necessity of locating the locally ambiguous DS [Deep Structure] position of a *wh*-word *on-line* ...In cases of LF movement the parser simply need not fill a gap and consequently no island effects result" (334).

In the following example of adjunct extraction (Pritchett's no 29), it is proposed that its unacceptability derives from unprocessability rather than ungrammaticality:

(35) *What did you eat_____ after John ruined e?

"eat" has an optional internal argument, so the chain headed by "what" can be assigned a θ-role and Case at this point, but after "ruined" has been encountered and identified as having an *obligatory* internal argument, the variable must be reanalysed as within this θ-domain - there is no other candidate. The fact that "ruin" and its θ-domain is part of a VP adjunct and not dominated by 'eat' means that the Theta Reanalysis Constraint is violated. For this reason, the sentence is unacceptable. (36) (Pritchett's no.30) also requires reanalysis of the variable, as the external argument of "ruined" rather than the internal argument of "believe"

but it remains within the θ-domain of the matrix verb as the VP complement, so no violation occurs:

(36) Who do you believe_____ e ruined the toast?

However, as Pritchett notes, an intransitive matrix verb, as in (37)

(37) *What did you sleep after John cooked_____ e?

avoids the need for reanalysis and hence no violation of the TRC can take place. The additional requirement, that the search for a gap be head-driven, rescues this case, as the adjunct is not licensed by a local head - so "it is impossible to postulate gaps within an adverbial clause"(340).

Saah and Goodluck (1995) found that the TRC did indeed hold for their adult Akan speakers, but simply as a processing principle, and in non-reflective tasks, not necessarily in all experimental conditions.

Empirical studies of the processing of island constraints do not definitively confirm that processing is carried out locally, on-line, rather than globally, off-line. However, the semantic opacity of ungrammatical items may also affect test performance, and the conclusions that may be drawn from test results. Crain and Fodor (1987) showed that in semtence matching tests, what made matching more difficult was "correctability": regardless of the cause of its unacceptability a "correctable" (ie interpretable) sentence took longer to process than an "uncorrectable" one. In such tests a sentence is presented on a screen, and two seconds later a second sentence appears below it. The subject has to judge whether or not the two sentences are the same, and the decision time is measured. In the case Crain and Fodor discuss, Friedman and Forster (1985) argue that core grammar or UG violations such as Subjacency are "overgenerated" by the grammar - are perceived as well-formed and do not therefore take longer to process than their grammatical mates (ie matching takes less time), while language-specific rule violations (eg subject-verb number agreeement) do not enjoy this advantage. Crain and Fodor (1987) demonstrate that the evidence does not warrant such

a conclusion and that in this case processing difficulty is not related to the grammar.

One later contribution to the sentence matching debate (Forster and Stevenson 1987) suggests that people judge constraint violations according to principles of "local well-formedness"- they will prefer sentences with locally well-formed phrases to those with locally ill-formed phrases, even if the global effect is implausible or ungrammatical. Thus (38) is locally ungrammatical (from Forster and Stevenson 1987):

(38) *John disappeared Mary during the party.

because an intransitive verb is followed by a NP in object position. (39) on the other hand, like many Subjacency violations, is locally well-formed – each clause is well-formed:

(39) *Who did Harry believe that John liked Bill?

Hence, performance on such sentences reflect parsing processes, and no particular grammatical theory need be invoked, nor need distinctions between "correctable" and "uncorrectable" violations be drawn.

This conclusion poses the question of the relationship between grammar and parser: if the latter is primarily affected, and misled by local conditions, it must be operating at a different level of representation from the grammar, even if it uses the grammar to identify the violations as such: "no algorithm for direct application of linguistic constraints (the competence grammar) will totally account for parsing stages" (Saah and Goodluck 1995: 406).

One final issue concerns the influence of the L1 grammar in parsing L2 sentences. Cook (1989), quoted in Cook (1990) found that the time taken to make grammaticality judgements correlated with *judgements,* not with parsing or processing difficulty. Cook (1990:594) reports that shorter sentences in his binding experiment were not easier to process than longer ones, but that the L1 parameter setting seemed to determine the speed of decision-making.

One interpretation is that language processing reflects the settings for the governing category parameter that are still latent in the mind; difficulty increases as the possibility that the sentence could have ambiguous meanings for different settings increases. (594)

Cook (1990:592) also found that advanced L2 learners performed at native levels in the timed comprehension test but that they took longer to do so. He then goes on to suggest that parameters are integrated with parsing, that UG in fact is "a parser that contains the principles of word order but continuously sets and resets word order parameters from evidence as it processes language" (594). If this is the case, then it can be argued that "difficulty" is not invariably a straightforward function of length or complexity. Difficulty is a feature of some syntactic structures that, depending on the relationship between the parametric settings of different languages, affects processing not always in a linear way.

To summarise:

(1) Knowledge of wh-movement rules in a language entails knowledge that a wh-operator heads a chain (or that a filler requires a gap, in GPSG[12] terminology);

(2) In cases of ambiguous or ill-formed sentences, people will incorrectly hypothesize a gap as soon as the grammar and semantic representations allow it;

(3) When forced to continue a parse, people will continue to attempt to complete it as soon as possible, through successive gap reanalysis, and will prefer locally well-formed phrases to locally ill-formed phrases, even in ungrammatical sentences.

4.9   Lexically-based Accounts of Wh-Movement Constraints

Not all the facts regarding constraints on wh-movement can be explained by syntactic theory. Chomsky (1986) refers to lexical choice as affecting the island status of extraction from NP complements. Other writers have argued that non-structural factors, such as discourse (Erteschik-Shir 1973) or pragmatic features can account fully for island effects. In this section we discuss some of this work.

We begin with the case of NP complements. There are clear differences in acceptability between the following:

(40a) What did you say she stole?

(40b) *What did you mutter she stole?

(40c) ? What did you hear the news that she stole?

(40d) *What did you hear a whisper that she stole?

The difference between (40a) and (40b) lies in the choice of matrix verb. "Mutter" is a lower frequency verb than "say", and although not ungrammatical in this context, is much less familiar. "Mutter" also has a much more specific meaning and gives additional information about manner of speaking. It would appear that some verbs do not allow extractions from their complements as readily as others, and that this feature is less to do with their subcategorization rules than with semantic issues of specificity and referentiality. Kluender (1992) argues that focussing on the manner causes a presupposition of factivity, which blocks extraction. (40c) and (40d) show that nouns can also affect acceptability in this way. Similarly, non-bridging verbs like "know" and "realise" (Erteschik-Shir 1973) or NPs that presuppose their complements are fact block extraction of adjuncts:

(41a) When$_i$ did you know$_i$ she stole the dress$_{-i}$?

(41b) When$_i$ did you hear the rumour$_i$ that she stole the dress$_i$?

In (41a) the presupposition is that stealing the dress is a fact and "when" can be moved from the upper clause, but not the lower/embedded clause. In (41b) stealing the dress is not necessarily a fact and "when" can refer either to hearing the rumour or to stealing the dress.

Kluender (1992) also develops a more elaborated argument for principles of predication, based on the specificity of the moved element: the more referentially specific a moved NP, the more acceptable it becomes. Hence (42a) is preferred to (42b):

(42a) Which article don't you remember who wrote?

(42b) What don't you remember who wrote?

Cresti (1995) notes, by the way, that a "wh X" phrase like "which article" is semantically richer and has more descriptive content, so is generally easier to interpret, especially in decontextualised sentences.

Referential specificity in Kluender's argument correlates highly with low frequency and with open class membership. Existential verbs and even unaccusatives correlate more highly with referentially specific NPs than do activity verbs, leading to the claim that processing constraints require a prominent, or foregrounded, component paired with a backgrounded component. Referentially specific complex NPs are most acceptably predicated of semantically light verbs, such as statives, as in "hold the belief that..." (Kluender's Predication Principle II). In long wh-movement however, referentially specific NPs are easier to process than pronouns in initial argument position. Kluender (1992) extends this argument to wh-islands: in short movement, closed class, non-specific pronouns are more successfully extracted from an island than specific elements, while the reverse is true for long-distance movement:

Which book did you wonder who bought? > What did you wonder who bought? >What did you wonder which man bought? > Which book did you wonder which man bought?

Wh-island complementizers degrade as they become more specific (0 > that > wh). Hence Kluender's account provides a metric for unacceptability for wh-movement sentences, deriving from combinations of increasingly specific matrix verbs, complementizers and NP complements, together with decreasingly specific wh-elements in initial position. ERP studies confirm that closed class, high frequency items are easier to process - that is, evoke greater negativity - in many different positions in sentences.

Kluender and Kutas (1993) develop this approach further: the Subjacency condition can be reduced to a processing constraint determined by short-term memory capacity. On a par with the difficulty in parsing grammatical multiply embedded sentences, the processor is unable to hold an extracted element in memory over a second wh-element. In support of this hypothesis, their experimental work with ERPs points to reductions in acceptability (= increases in electrical activity) according to type of complementizer (that > if > wh-)

regardless of grammaticality. There remain unresolved problems with this approach. Kluender and Kutas seem to equate processing ease with acceptability, and acceptability with grammaticality. Although referential specificity slows down processing, and multiple violations of the Predication Principle do, eventually, result in ungrammaticality, there are clearly other routes to ungrammaticality which do not engage with the issue of more or less specificity.

Work on processing garden path sentences indicates that the processor will prefer the "easy option" and will resist backtracking or reanalysis till the last moment, but that processing difficulty *per se* does not cause breakdown. Culicover and Nowak (1995) have also pointed out that island constraints represent an accumulation of violations of acceptability caused by lexical unfamiliarity rather than syntactic constraints and that "if a relatively complex construction is experienced frequently enough, it will be as easy to process as a simpler construction, other things being equal" (40).

Whether this argument holds up cross-linguistically, in the way that the syntactic account does, remains to be seen.

<div align="center">*</div>

---

[1]In Italian, 3rd person subjects must be licensed by topic antecedents (otherwise they are uninterpretable) (Grimshaw and Samek Lodovici (1998)) . In English, too, null subjects are permitted in certain registers, when a discourse antecedent makes them interpretable.

[2.] The Minimalist Program (Chomsky 1995) generalizes the ECP as a "descriptive cover term for various kinds of violations that are marked at LF, among them violations of the economy principle (Relativized Minimality)" (91). However, although Subjacency violations also fail this principle, they are not present at LF.

[3.] Theta marking describes assignation by a verb, or other predicate of a thematic role, such as agent or theme, to a constituent, according to the theta criterion, by which each argument is assigned only one theta role, and each theta role is assigned to only one argument (Haegeman 1994: 71-72)

[4] It should be noted that the theoretical basis for the ban on extracting from an adjunct clause (not L-marked and therefore a barrier) has been questioned: Lightfoot and Weinberg (1988) have argued that adjunction to PP, analogous to adjunction to VP, would avoid the crossing of the barrier; in addition, the main clause IP can only act as a barrier if movement from the adjunct clause takes place before the main clause auxiliary is moved into CP – hence, an island effect seems to require rule-ordering, which is dropped from the theory in "Barriers".

[5].Rizzi (1982:...suggests this node may be either VP or IP, but does not investigate the question further.

[6] Rizzi (1982) atrtributes the marginality of this sentence to its complexity rather than to any possible Subjacency infringement

[7] In languages where the verb inflects for tense and number, IP can head-govern the subject through T (the maximal projection of TENSE) according to the Split-INFL Hypothesis (Pollock 1989)

[8] Cook (1990) argues "that the alternative settings for the two-valued pro-drop parameter are latent in native English speakers in that they produce null-subject sentences in certain registers"(594)

[9] Otsu's test items have been criticised for syntactic complexity, while the questions asked by Goodluck et al were straightforward and short and perhaps tested the child subjects' knowledge of wh-movement constraints more precisely.

[10] Uziel used the Michigan Test of English Language Proficiency, a placement test for foreign students intending to study at US universities. A multiple choice test of grammar, vocabulary and reading, it seems less than appropriate as an instrument for assessing subjects in this kind of experiment (Spolsky 1995).

[11] There is also evidence that adults, like children, prefer "flat" constructions like conjunctions to "tall, thin" ones like relative clauses (Tavakolian 1981)

[12] Generalized Phrase Structure Grammar, a non-transformational theory which focusses on developing phrase structure.

CHAPTER 5

RESEARCH HYPOTHESES

5.1 **Summary of Critical Period and Age Research**

5.2 **Summary of Ultimate Attainment Research**

5.3 **Summary of Wh-movement Constraints**

5.4 **Research Hypotheses**

5.1   Summary of Critical Period and Age Research

To date, and possibly for a long time to come, there has been no definitive neurological evidence that language acquisition is determined by the maturation of a "language function" in the brain On the other hand, complete post-pubescent acquisition of a first language is unknown, although evidence from acquisition of signing amongst deaf individuals suggests that not all levels of language are inaccessible in late first language acquisition.

In second language acquisition, children have a clear advantage over adults in all language areas. Native-like attainment by children is not inevitable, but neither is non-native like attainment by adults. However, adult attainment is considerably more variable, and much more likely to fossilize at an intermediate stage, short of adult native speaker attainment. For second language acquisition, therefore, child learning seems favoured over adult learning. Possible causes of differential outcomes are neurological changes in later life. While the language faculty itself remains intact, short-term memory limitations from middle age onwards inhibit comprehension and processing in the first language, and as this kind of memory is strongly implicated in second language acquisition, older adult learners may be disadvantaged. For older and middle-aged bilinguals there may too be a tendency to mix grammars and therefore to be less rigorous in disallowing properties of the L1 which are ungrammatical in the L2.

Questions that still require to be addressed are:
* in what way does the mental representation of the L2 grammar held by adult learners differ from that held by native speakers on the one hand and child learners on the other?
 * what evidence is there of impaired language learning abilities in later life?

5.2   Summary of Ultimate Attainment Research

The literature provides many examples of near native proficiency and near-native competence amongst adult learners. The nature of near-native competence is more controversial. If the potential for native-like competence hinges on access to UG, it should

fall out that knowledge of properties of UG will be diagnostic of complete, incomplete or divergent competence, relative to native competence. Ultimate attainment studies do not agree that UG properties are privileged in the L2 grammar. Even where UG seems to be accessed, certain aspects of a principle may be denied such access: fossilization can set in at *any* level of acquisition, and may indeed be inevitable.

Questions that still require to be addressed are:
* If fossilization is inevitable, at what point do highly proficient learners fossilize?
* Is UG "hierarchically available" to L2 learners such that completely target-like acquisition of UG-based properties of the L2 is not possible for all, and is this access determined by L1 knowledge?

5.3   Summary of Wh-movement constraints

Subjacency and the ECP are selected for this study because they appear to be well adapted to deal with the research questions raised above. They have already been used in a wide variety of SLA studies concerned with age of exposure and ultimate attainment. The Subjacency Condition exhibits both parametric variation and hierarchic features (weak and strong violations) and so can be used to probe intuitions at a greater degree of delicacy. It also allows the construction of long distance movements items (ie with multiple embeddings) suitable for testing the intuitions of advanced learners.

Questions that still require to be addressed are:
* Do Italian speakers of English accept Italian settings of the Subjacency Condition in English, where they are unacceptable?
* Do highly proficient and near native Italian speakers of English observe the same hierarchy of acceptability as native speakers?
*   Are adult learners more inclined to give divergent judgements than child learners?

5.4   Research Hypotheses

The hypotheses listed below attempt to operationalise the research questions summarised in 5.1-5.3. The primary concern of the experimental study undertaken for this dissertation is with the limits of second language acquisition in adulthood. This requires two comparisons to be made: the first between adult learners and child learners, and a second amongst adults who arrived at different ages. We have seen that while several studies have investigated the former relationship, only one, Birdsong (1992), has produced evidence of the superiority of earlier over later adult arrivers. We attempted to obtain more robust data, by comparing groups of earlier and later adult arrivers, and by testing subjects on a single property of UG, which has the added advantage, as outlined above, of offering a rich source of parametric variation.

The second dimension of the study deals with the nature of near-native competence. Where White and Genesee (1997) found no difference between native and non-native subjects (from a variety of L1s) on Subjacency violations, after controlling for proficiency, Birdsong (1992), Sorace (1993) and Ioup et al (1994) found precisely that parameterized properties of UG did evoke both divergent and indeterminate responses. We propose here to add to this body of research into ultimate attainment by testing speakers from a single L1, Italian, on a property of UG which varies parametrically between Italian and English and which includes weaker and stronger types of violation. This investigation allows a comparison of the acceptability hierarchy derived from both native speaker and non-native speaker respondents with the grammaticality hierarchy derived from the theory.

The subjects tested with a Grammaticality Judgement test were Italian speakers of English who had been resident in Britain - mostly in Scotland - for a minimum of 5 years, and whose proficiency in English was estimated to be at or near native speaker level, for fluency, grammar and vocabulary. Details of the selection procedure are provided in Chapter 7.

5.4.1   Age of Arrival Hypotheses
a. In a Grammaticality Judgement test of Subjacency and the ECP scores of Italian speakers of English who arrived before the age of 15 will not differ significantly from those of native speakers, but will differ from those of adult arrivers.

b. Age of arrival in adulthood will correlate negatively with scores on grammatical and ungrammatical wh-movement sentences, such that the later the age of arrival the greater the divergence from native speaker scores.

5.4.2   Acceptability hierarchy hypotheses

a. Native speaker subjects will show sensitivity to degrees of ungrammaticality according to the "Barriers" (Chomsky 1986) account. ECP violations will be rated lower than Subjacency violations, and some types of the latter will be rated lower than others, depending on the number of barriers crossed. The order of acceptability will be as follows:

Table 5.1.   Acceptability Hierarchy for Wh-movement Constraint Violations

Violation Type	Number of Barriers Crossed	Sentence Type	Extraction Type	Identity Number	Acceptability level
Subjacency - parameterized	1	wh-island	object from finite clause	1	most acceptable
Subjacency - parameterized		wh-island	object from nonfinite clause	2	
Subjacency - invariant		wh-island	object	3	
		noun complement	object	4	
			adjunct	5	
	2	relative clause	object	6	
		adjunct clause	object	7	
		wh-island	subject	8	
ECP		wh-island	subject	9	
		noun complement	subject	10	
		relative clause	subject	11	
		wh-island	adjunct	12	
		noun complement	adjunct	13	least
		adjunct clause	adjunct	14	acceptable

In addition, extractions of objects will be rated as more acceptable than extractions of subjects.

b. Italian speakers of English who arrived before the age of 15 will obey the same hierarchy as native speakers.

c. Adults will observe the hierarchy as far as strong violations and ECP violations are concerned, but the later the age of arrival in adulthood, the more indeterminate the judgements of weaker Subjacency violations.

d. Adult arrivers will tend to rate Subjacency violations which are grammatical in Italian but not in English higher than native speakers or child arrivers, and ratings of these violation types will correlate negatively with age of arrival.

5.4.3   Additional hypotheses

a. The grammar of an L2 speaker of English who has been resident continuously in the L2 environment for a minimum of 5 years can be deemed to have completed its development. Length of residence of these subjects will therefore not predict test performance.

b. Although not all such speakers will have achieved native levels of performance proficiency in grammar, phonology or lexis, they can still have nativelike intuitions about acceptable and unacceptable sentences in the target language. Measures of proficiency above a certain level will therefore not necessarily correlate with test performance.

c. As type of input as well as amount of exposure may have a bearing on the final state of the L2 grammar, subjects who had formal instruction in English (assuming a focus on form), and those who use English as their main language at home and at work will perform closer to native speaker norms than those who acquired English informally and whose main language of use is Italian.

d. As age has been shown to correlate negatively with processing ability, subjects over 50 are likely to rate complex grammatical sentences lower than younger subjects. Subjects over 50 may also show more flexibility in responses to Italian settings of the Subjacency Condition, by a greater tendency to accept them.

*

CHAPTER 6

EXPERIMENTAL METHODOLOGY

6.1   Introduction

6.2   Grammaticality Judgement Tests in Language Acquisition Research

6.3   A Critique of GJ Tests in SLA Research

6.4   The Design of the Test

6.5   The Pilot Studies

6.1   Introduction

This chapter describes the rationale for the choice of experimental methodology. We recount the short but controversial history of grammaticality judgement (GJ) tests in Second Language Acquisition research; discuss some of the changes that have been adopted to improve their validity and reliability; describe how GJ tests have been used in a number of wh-movement studies; and seek to justify the particular form and content of the GJ test selected for the present study. Two pilot studies, carried out to test the materials and assist in the selection of an appropriate rating method are then described. We explain and provide a rationale for the adoption of the technique of Magnitude Estimation. Finally, the criteria for the selection of subjects and the test procedure for the experiment are explained.

6.2   Grammaticality Judgement Tests in Language Acquisition Research

Most studies of both first and second language acquisition within the framework of the Principles and Parameters model have used the technique of the GJ test to probe the intuitions of speakers, both native and non-native, about the target grammar. Such tests generally involve asking linguistically naive respondents to distinguish grammatical from ungrammatical sentences, usually presented without any context and in a way that requires the respondents to focus on syntax rather than semantics or pragmatic features. The task thus probes metalinguistic intuitions to discover whether, or to what extent, such intuitions match what is grammatically possible and not possible in the target language.

There are many different procedures for GJ tests: for adults, the test items may be presented in a booklet or on a computer screen or recorded on audio tape; the respondents may have a set time in which to make their judgement (paced tests), or their response time may be measured (timed tests); they may be required to respond with a simple "yes/no" (dichotomously), or rank their response, according to the perceived degree of grammaticality or acceptability. For children, sometimes as young as 2, a number of rather ingenious tests have been devised, including toy-moving and picture choosing tasks, which are comprehension rather than the metalinguistic tasks usually devised for adult subjects, but

which, it is claimed, also tap into the child's mental representation of the grammar. Modifications of the "basic" GJ test have included asking respondents to choose which one in a pair of sentences they prefer. Experimenters have also had subjects correct those sentences they rate as ungrammatical, and reflect on their judgement in discussion with the experimenter, on tape, or by indicating how confident they felt in each case.

Linguists have made use of their own and other linguists' intuitive knowledge of the grammaticality of sentences for many years. Generative grammarians in particular have relied on their own, trained judgements to distinguish between grammatical and ungrammatical sentences. Grammaticality is, as Haegeman explains, a "theoretical notion" (1993:7): "A sentence is grammatical if it is formed according to the grammar of English as formulated by the linguist". All native speakers, on the other hand, are able to make intuitive, non-theoretical judgements as to the *acceptability* of sentences, including sentences they have never seen before. They can also rank sentences in order of acceptability. Not only is there usually a very high degree of agreement among native speakers on grammatical and ungrammatical sentences, there is also a high correlation between grammarians' judgements and native speaker informants' judgements - thus between grammaticality and acceptability (Newmeyer 1983). Since Bloomfield (1933) at least, NS intuitions have "constituted primary grounds for the inclusion of structures in the description of languages" (Chaudron 1983:343). Chomsky (1957,1965) insisted that if the task of linguistic theory is to characterise the competence of the native speaker, then the theory can only be tested by evidence from the speaker's competence.

The precise extent to which GJ testing can actually provide such evidence has been continuously debated over the last 30 years. It is generally recognised that production data is too "noisy" and incomplete to provide a reliable guide to a speaker's competence: performance factors such as attention, memory, affective and psychological variables will affect individual performance. Variability (Tarone 1988) and avoidance strategies are especially likely to affect second language speakers' production. Grammaticality judgement tests are nonetheless admitted by most researchers to elicit another kind of performance. "They are not a direct reflection of competence, for competence is an abstraction" (Gass 1994:306). "Judgements of acceptability ... may fail to provide direct evidence as to grammatical status because of the intrusion of numerous other factors" (Chomsky 1986b:36). They too are affected by variation in attention and memory, fatigue, inhibitions caused by the

experimental setting, and other non-linguistic factors (see below). Perfectly grammatical sentences may for example prove too hard to process (as in the notorious case: "the cat the dog the man hit chased ran") or prescriptive, "school grammar" rules of good and bad grammar will overrule intuitions. Most respondents can also, given enough time, find a context which renders the oddest sentence acceptable. Levelt (1972) attacks this kind of investigation: not only is there "a complete absence of arguments in the literature in favor of the thesis that linguistic intuitions reveal the underlying linguistic competence" (Levelt 1972:23); such intuitions are also "very derived and rather artificial psycholinguistic phenomena" which develop late and are heavily influenced by explicit instruction. (Levelt et al 1977:88). Levelt (1972) describes an informal experiment where sentences from papers on transformational linguistics showed a wide variety in acceptability when judged by linguists out of context (ie when matching grammatical items were lacking).

Grammaticality judgement tests have however proved of enormous value in linguistic and applied linguistic research, since they seem to produce data which are *closer* to matching speakers' mental representations of a language than other data, and which stand up to a variety of triangulation procedures. Moreover, unlike production tasks, GJ tests can present subjects with sentences they have never seen or uttered before, and successfully elicit confident judgements, which suggests that speakers' competence is *underdetermined* by the input they receive. GJ tests also tend to be confirmed rather than contradicted by other evidence where available. Thus, in spite of their limitations, data collected by this technique can provide useful insights into informants' grammatical knowledge, so long as the tests themselves are designed to control for as many of the interfering factors as possible (Birdsong 1989). Carroll, Bever and Pollack (1981), in a much-quoted critique of the use of linguistic intuitions, conclude nevertheless that, while manipulating the mental state of subjects in experimental conditions can alter linguistic intuitions, the differences are systematic: "this raises the possibility that there are relatively stable mental schemata that underlie manifest intuitions of similarity" (379). Linguistic intuitions are therefore at the same time primitive manifestations of linguistic knowledge and "complex behavioral performances".

In the field of second language acquisition research, it has been argued that the very instability and increased variability of L2 speakers' intuitions about the target grammar adds to the value of studying them. Chaudron (1983) points out that "the study of second language

acquisition has raised the unique issue of how source and target language grammars interact in the development of an interlanguage (IL) and what this interaction indicates about the structure and operation of linguistic awareness itself"(346). Metalinguistic data can provide insights into both the IL grammar and its development. In the case of GJ testing of ultimate attainment, an individual's grammar is - we can assume - at a final, stable state: the object of investigation then becomes this grammar as compared to those of other L2 ultimate attainers and to the NS grammar.

Acceptability judgements have continued to be used in both L1A and L2A studies because (a) in the main, they have been consistent *enough* with performance data to give them credence and (b) no more valid and reliable procedure for tapping linguistic competence has been found. (Sorace 1990).

In line with most of the literature, we shall term such tests Grammaticality Judgement or GJ tests, while acknowledging that they involve judgements of *acceptability*.

In the case of the present study and of previous research using Subjacency and ECP violations of wh-movement, GJ tests are particularly appropriate for a number of reasons. Firstly, unlike parameterised features such as Null Subject, the Binding Principle or adverb placement, knowledge or otherwise of Subjacency and the ECP tends not to appear in production data. Comprehension tasks, such as those used to some effect by Goodluck and her colleagues, are only appropriate where the violation is both interpretable and plausible, as in Subjacency violations in adjunct clauses, not where the violation is uninterpretable, as in ECP violations. Finally, GJ tests allow the experimenter not only to explore subjects' intuitions as to the acceptability or unacceptability of items, but also to discover to what extent subjects are able to rank different violations in a *hierarchy* of acceptability: this involves metalinguistic probing of a complex set of intuitions.

6.3   A Critique of GJ Tests in SLA Research

GJ tests have been used extensively in Second Language Acquisition research. At the same

time, the use of this technique for describing interlanguage has been equally extensively criticised. The arguments raised against the reliability of acceptability judgements in SLA studies have been persuasive (see discussion in Chaudron 1983:343-4 and Birdsong 1989), but have also served to spur researchers on to improve both their reliability and their validity in many respects. If such judgements are a kind of performance data, they will vary in accordance with a variety of known and unknown behavioural and situational factors. In recent years, researchers have been at pains to identify and control for these factors.

Ellis (1991) points to deficiencies in reliability, for instance the huge between- and within-subject variation often found: a subject may judge the same item differently within a single test and weak correlations between two trials of the same test have been found. Schmidt and McCreary (1977) however compared oral production data with respondents' ratings of written standard and non-standard data. Although they found non-native speakers more consistent than native speakers, both groups were found to be predictable in their responses. A strong correlation between results from two trials of the same test, a week apart was also found by Gass (1994): only syntactic features which her subjects were unsure of (i.e. had not fully acquired) elicited varying judgements (see also Johnson et al 1996). Ellis (1991) further argues that it is often impossible to determine the criteria respondents base their judgements on: are they really judging the grammar (the syntax) or the semantic or pragmatic plausibility, or even the discourse context of sentences?

There would appear to be several ways of lending GJ tests greater validity and greater reliability. Reliability has been improved by demanding more stringent selection procedures, such as matching respondents more closely for L2 proficiency, L1, age, socio-economic and educational background. Controlling for order of presentation of items, as suggested by Chaudron (1983:367), for example by continuous randomisation to avoid effects of context and fatigue has helped reliability as well. Replication, either with the same subjects at a later date or with another group at the same time, although standard procedure in much scientific and psychological work, is rare in applied linguistics. This is partly because of practical difficulties in finding suitable subjects, but also because even for apparently asymptotic learners it cannot be guaranteed that there has been no change in their L2 grammar by the time of a second test. Subjects may learn from the first test and apply that learning to the second test. Johnson et al (1996) found that their adult learners improved their scores significantly at the second test, taken three weeks after the first ($t = 4.36, p < 0.01$).

Several writers and researchers have called for the triangulation of GJ tests with other, more explicitly performance tasks, such as think-aloud protocols (Ellis 1991), a question-formation task (White and Genesee 1996), compositions (Phinney 1987) or correction of ungrammatical items (Zobl 1992). If subjects' performance in this kind of task is consistent with their performance on the GJ test, as regards the grammatical feature at issue, then it seems more likely that the experiment has in fact succeeded in tapping their grammatical knowledge:

> Ideally, performance data from various sources will converge in giving a picture of what abstract linguistic competence is like, and will allow one to make generalisations which do not depend on the results of a single test. (White and Genesee 1996:7)

If the purpose of an experiment is to prompt *intuitive* responses, rather than have subjects consciously try to puzzle out the source of the ungrammaticality of a test sentence, including tasks that require such conscious effort may in fact undermine the reliability of a GJ test. Although the nature of the experiment was completely different, it may be noted that the task set in Kluender and Kutas' (1993) ERP tests discussed in Chapter 4 , was designed to distract subject's attention away from the experimental purpose (subjects were merely asked to indicate whether a target probe word has appeared in the test sentence just presented).

Without a psychological theory of linguistic intuitions, the question, "What exactly are respondents judging?" cannot be properly answered. Moreover, in the case of L2 learners, even highly proficient speakers, it is probably harder to identify what criteria are being brought to bear in GJ tests. Ellis (1991), Goss et al (1994) and Davies and Kaplan (1998) report on studies of the strategies used by L2 speakers in GJ tests. In the most recent of these studies, Davies and Kaplan (1998) present the findings of group think-aloud protocols of native speakers of English learning French in the US at low intermediate level who were tested first on complex English sentences and then on simplex French sentences. Their subjects used very different strategies for judging in their two languages - employing translation, analogy and guessing and making far greater use of explicit, learned knowledge and semantic information for French. Davies and Kaplan point out that the use of different strategies may be a function of their subjects' low level of proficiency in the L2 but that this implies that GJ tests can only safely be used with advanced (asymptotic) learners. Gass

(1994) also suggests that indeterminacy led to guessing in her study involving a repeated test. However, Davies and Kaplan's categories (arrived at deductively) do not always convince: they place "feel" and "repair" in separate categories, although the discussions referred to under these labels appear equally intuitive; the most popular strategy for dealing with French sentences was 'learned' which the authors seem to apply to any cases where subjects *discuss* test items, rather than pronouncing on them immediately. Combining the categories "feel", "repair" and "learned" into one strategic category distinct from the strategies using meaning, translation etc reduces the difference between L2 and L1 judgements markedly. Research findings on the variability of L2 judgements in GJ tests seem so far inconclusive: there is little firm evidence that highly proficient L2 speakers use significantly different metalinguistic strategies.

Processing factors, including memory, attention and parsing, can be controlled to improve the validity of GJ tests, as will be discussed below. Greater attention to item selection allows for more realistic or authentic sentences that can be more easily processed than the "linguists' sentences" frequently resorted to.

In the present study it was decided to use a GJ test for the following reasons:

(1) Previous research in this area has generally made use of GJ tests, so comparisons can more readily be made;

(2) In spite of or perhaps because of the criticisms of these tests, they remain the best (least bad) and easiest way to explore speakers' intuitions about grammar;

(3) They can allow speakers to make fine distinctions between degrees of acceptability for constructions that occur rarely in production data;

(4) Experimenters can model L2 grammars, using a range of statistical procedures.

It was decided not to include production tasks, on the grounds that several studies have done without them (eg Martohardjono 1992, Uziel 1993), but also because it was considered impractical to require the kind of subjects envisaged to carry them out: their time would be limited, unlike that of captive university students.

6.4   The Design of the Test

The form of GJ finally settled on for the present study was an untimed, unpaced test of 82 written items, presented one by one on an OHP screen. Subjects were given 6-12 seconds to read and then rank each item on self-created interval scale, using Magnitude Estimation. The test lasted for approximately 45 minutes.

In this section, we discuss the reasons for selecting each of these features, referring to previous research in this area.

6.4.1 Modality

Most GJ tests in Second Language Acquisition have used written rather than aural presentations of materials. While aural presentations may appear to be more natural and therefore to elicit more valid judgements, they entail certain problems.

Both the Johnson and Newport studies (1989,1991), discussed extensively above, used aural presentations. Johnson (1992) replicated the earlier study with some of the same subjects and the same materials in a written version. She found that adult learners (although not the child learners) made more than twice as many errors in the aural version as in the written version of the test. Johnson concludes that "some source of grammatical knowledge is available to the adult learners during the written test that is not available to them during the auditory test" (Johnson 1992:233). She speculated that the mode of learning (by formal instruction) of the adult learners may have been responsible. However, neither this study nor Haig (1991), where subjects were found to be less accurate in rejecting ungrammatical wh-movement sentences in an aural test than in a written, timed their tests. In fact, Johnson (1992) stressed to subjects "that they should take all the time that they needed to complete [the test]" (224). Thus there is no evidence to show that it was not simply the time factor that made the difference.

In a recent investigation of the modality effect, Murphy (1997) presented English and French native and L2 speakers with aural or written versions of the same test. Both presentations were timed. She found that accuracy in judging grammatical sentences was not affected by modality, but that for ungrammatical sentences (Subjacency violations), highly proficient L2

speakers were significantly less accurate and slower in the aural modality. Moreover, even native speakers showed a significant modality effect in their reaction times. There was a difference, although not significant, between native speakers and L2 speakers on reaction times for ungrammatical sentences. Murphy concludes that using both modalities provides more compelling evidence than only using one, but her findings do not really support this: she offers no evidence that the aural modality gives a different perspective on judgements (as Johnson claims for adult learners), for example that accuracy rates differ on different types of sentence/violation, as between modalities. What her study does suggest is that aural presentations can add to processing load, requiring a longer response time, and producing less accurate responses, perhaps because of working memory deficits. It does not suggest that subjects tap into a different (e.g. reduced) set of intuitions when confronted by an auditory test, as Johnson (1992) proposes. Research into differences between listening and reading (in L1 processing) has demonstrated superior performance from subjects in visual conditions; and has shown that comprehension in listening takes longer than comprehension in reading (Murphy 1997).

The advantages of an aural modality are that aural language input is more common in everyday life, so an aural GJ test should be more naturalistic than a written one. Also the test automatically becomes paced. On the other hand, this mode of processing may place a heavy burden on working memory. It also compels the experimenter to keep items short, with a minimal number of embeddings. Complexity, however, is a significant factor in knowledge of wh-movement.

In the Subjacency test reported in White (1988) test materials were presented simultaneously on tape and in written form, but apparently only to ensure a paced test. As pacing was not an issue (see below) and for the reasons given above it was decided in the present study to have subjects read test items.

6.4.2 Length

Overall test length must depend in part on the type of subjects tested and the resources available. For example, Weber-Fox and Neville (1996) were able to conduct behavioural and ERP tests on 61 Chinese/English bilinguals with a GJ test of 240 sentences presented on computer screen one word at a time, followed by 4 standardised proficiency tests (of grammar and reading span), presumably lasting several hours. Most reported GJ tests

however have had between 60 and 100 sentences thus avoiding fatigue and attention factors. Although numbers do vary widely. Johnson and Newport (1989) had 180 sentences - which must have taken over 3 hours to judge - while Schachter 1989 had 66, and Uziel (1993) 100. For the present study, 86 sentences were prepared.

The test was neither paced nor timed. It was expected that the subjects to be recruited would include several individuals for whom the GJ test in itself would be an unfamiliar and perhaps unsettling experience, at least at first. Every effort was that made to ensure that the test would be as unthreatening and informal as possible. This entailed not using a computer (thus not measuring reading or judgement times accurately) and not dictating the pace of judging by using a timer or an audio tape. The study of very elderly L1 speakers by Pye et al (1990) described in Chapter 2 allowed subjects, some over 80, to complete the task at home, in their own time. In these conditions, albeit perilously uncontrolled (how many grandchildren tried to help?), subjects were willing to judge 192 written sentences of considerable length. In the present study, therefore, subjects were simply encouraged to read and judge as fast as they could, and not to ponder or try to correct.

More recent studies (e.g. Uziel 1993, Murphy 1997) have recognised the importance of controlling both for sentence length and for lexical content. Minimising differences among sentences, which may distract attention from syntactic features and also increase the processing burden, should improve the validity of results. Schachter (1989) had sentences ranging from 7 to 17 words, and admitted that this variation could have been a confounding factor. More recently, researchers have generally tried to keep items short and have tried to match them for length (White and Genesee 1996). Uziel (1993) (in an aurally presented test) kept sentences to between 9 and 14 syllables in length. However, length is not simply a validity question: longer sentences place a heavier burden on working memory, while tracking the grammaticality or otherwise of a sentence through a series of embedded clauses is more taxing of both processing and syntactic knowledge. Hence, in studies of native speakers or of ultimate attainment in L2, using long and complex sentences as well as shorter items may improve the reliability of results. Pye et al. (1992) used exceptionally long sentences[1]: a length effect (statistically significant) was found for all age groups, including the control group of college students, for grammatical sentences (364-5). Yet this did not affect accuracy in distinguishing between grammatical and ungrammatical items, except in the case of the oldest adults (70 and 80 year-olds). The question arises: in the case of lowered

accuracy, how to separate processing difficulty from deficits in grammatical knowledge? The studies of processing discussed in Chapter 4 attempted to answer just this question. We can conclude that while processing difficulty caused by length of sentence or utterance, number of embeddings, lexical factors such as frequency or referentiality have been shown to play a part, we cannot yet be confident in determining the extent of their effect. In addition, difficulty may be a function of the relationship between parameter settings in the L1 and the L2.

To sum up, varying test items systematically according to length and number of embeddings could give more opportunity to identify and isolate processing and parsing difficulties. For highly proficient NNSs, and for NSs, longer sentences should avoid the problem identified in J&N (1989,1991) of making test items too easy to categorise as grammatical or ungrammatical. Until the psychological processes involved in acceptability judgements are further investigated and better understood, an approach which presents a variety of types of sentence may be optimal. For the present study, sentences were held at 9,11 or 14 words. The number of embedded finite and non-finite clauses varied between 1 and 2.

Following Levelt (1972), every effort was made to ensure that the sentences should not be lexically or semantically strange or unnecessarily complex. In addition each grammatical control item was lexically matched to a violation item[2]. Lexical matching has also been a characteristic of more recent Subjacency studies (Martohardjono 1992, Uziel 1993, Murphy 1997), although it is rarely possible to produce exact matches where only the violating word differs between grammatical and ungrammatical versions. For example, in an ERP (event-related brain potential) experiment with Subjacency violations, McKinnon and Osterhout (1996) were able to achieve this, but at a cost: they had only one type of Subjacency violation (extraction of object from temporal finite adjunct clause), and they were obliged to devise some rather unusual grammatical wh-movement controls (e.g. "I wonder which of his staff members the candidate was annoyed that his son was questioned by"). With NS subjects, they obtained very low accuracy rates (48%) on such items, which may have been a function of the oddness of the sentences.

6.4.3   Rating and Ranking methods

Chaudron's (1983) review of acquisition studies using metalinguistic judgement procedures lists a very wide variety of ranking measures. In child FLA studies, measures tended to be

dichotomous or nominal ("wrong/right", "good/silly") (356). In adult SLA studies, nominal scales are still the commonest, but 3- to 10-point ranking scales have also been employed. Chaudron notes that rating scales in psychometric testing produce more reliable results as they increase towards 20 points, although it is maintained elsewhere (Nunnally 1978) that reliability increases rapidly up to 7 points, but very little after 11. Yet it is remarkable how many SLA studies maintain either dichotomous rating or very short scales. The former are more readily adapted for use in statistical procedures and may be more straightforward to administer especially where L2 subjects are not highly proficient, but they can hide subjects' uncertainty about the status of items. It is not uncommon for experimenters to instruct subjects to class items they are unsure of as ungrammatical. Bley-Vroman et al (1988) included "not sure", but lumped these responses together with 'impossible', for no principled reason (Gass 1994:310-11). Coppieters (1987) counted "not sure" as correct! Uziel's replication of Martohardjono (1992) included "not sure" and "don't understand" but marked them as incorrect when scoring (Uziel 1993:71). It can be argued however, as Sorace (1990) does, that inclusion of a "don't know" or "not sure" category leads only to confusion. Sorace (1990) points out that many learners - even highly proficient ones - are temperamentally inclined not to commit themselves to definite responses; moreover "not sure" might include both responses where the learner is unsure of the status of an item and responses where she believes this status is indeterminate (eg subject to dialectal or stylistic restrictions).

Ordinal scales, with more than two scale points, allow respondents to make finer-tuned and perhaps more accurate judgements and can hence improve reliability. With such scales, numbers or letters tagged with phrases indicating gradations of acceptability have been used. Schachter and Yip (1990) used a 4-point scale labelled "clearly grammatical", "probably grammatical", "probably ungrammatical" and "clearly ungrammatical". They had no category for "not sure" and Gass (1994) asks "how do we interpret the middle two categories?" Birdsong (1992), with L2 speakers of French, had "not at all acceptable - I would not say it" - "acceptable in rare contexts" - "acceptable in about half the contexts" - "acceptable in most contexts" - "completely acceptable; I would say it". Although the longer scale allows a wider range of judgements, the merging of acceptability with idiolect suggested by the descriptions for the first and last points seems unwarranted with non-native speakers.

Chaudron (1983) urges researchers using GJ tests to be "as explicit as possible in describing

the task to the respondents" (368), in view of the apparent variety of criteria used by respondents in making judgements. Many of the studies surveyed here display, as far as one can tell from the published reports, a lack of rigour and clarity in the instructions given to respondents. Johnson and Newport (1989) for instance, state that respondents were "instructed to ... make a judgement as to whether or not the sentence is a grammatical sentence in English, circling yes if they thought the sentences were fine (sic) and no if they thought it contained an error of some sort."(230). While phrasing instructions to help subjects focus on grammaticality rather than semantic or pragmatic features is clearly essential in GJ tests, many experimenters have provided terms for nominal judgements such as "possible/impossible in English", "good/bad" and "correct/ incorrect" which seem to invite prescriptive rather than descriptive judgements. Pye et al's (1992) study of elderly native speakers rather perversely - given that older people are more likely to make prescriptive judgements - employed the terms "ungrammatical, bad English" and "grammatical, good English" for either end of their 7-point scale.

As this brief summary reveals, several studies have used ordinal scales, thus perhaps improving the face validity of their experiment, but have then failed to maintain the distinctions elicited by the use of such scales in the statistical procedures which followed, sometimes arguing that the numbers of intermediate or indeterminate responses were very small (see Bley-Vroman et al 1988). Thus, valuable data may be lost. On the other hand, the fact that ordinal or interval scales are set by the experimenter, and described according to his or her assumptions about acceptability may, even with carefully worded instructions to subjects as to what to take into consideration and what to ignore when making judgements, run counter to subjects' own sets of assumptions or intuitions. For example, a scale that includes no point for "not sure" or "don't know" forces subjects who are not sure to guess more or less randomly; a scale point labelled "acceptable in about half the contexts" would tax the judgement even of an experienced linguist.

Interval scales permit much finer discriminations between sentences, or between different degrees of (un)grammaticality by measuring the differences between them based on a metric marked out in equal intervals. Thus respondents can specify exactly how far from a baseline standard of - say - grammaticality a particular sentence should be placed, then place a second sentence a certain number of intervals from the first. A ratio scale gives measured proportions between judgements. White (1988) had subjects mark a 9 cm line labelled

"correct" and "incorrect" at either end - a kind of ratio scale - allowing them to make their own proportional judgments, but there are few if any studies that use such scales.

6.4.4 Magnitude Estimation

The problem of reliability of judgement first raised by Levelt (1972) concerns not only the best way to present sentences in order to elicit "uncontaminated" judgements but also wider theoretical issues. Levelt (1972) noted that "a sentence which looks grammatical in isolation may nevertheless look ungrammatical if compared with other sentences" (25). His solution is to elicit ranking judgements, rather than absolute judgements, about degrees of grammaticality among different sentences, displaying different rule violations[3].

The technique of Magnitude Estimation (ME), borrowed from the field of psychophysics and used quite frequently in studies of speech perception (see below) has recently been employed in SLA studies to enable subjects to produce interval scales of grammaticality (Bard, Robertson and Sorace 1996). The technique allows subjects to make their own interval scales by assigning numerical or other scalable values, such as line lengths, to a series of stimuli. Devised to measure subjective estimates of sensory physical phenomena such as brightness, loudness or weight more accurately than ordinal scales, it was later used in psychosociological attitude studies in the 1960s. The subject is shown a stimulus or modulus demonstrating the target property (e.g. brightness, or a political opinion) with a number associated with it; the subject then responds to succeeding stimuli by allotting numbers that reflect the proportionate increase or decrease in the property. Thus if the modulus is allotted the number 10 a stimulus that feels 10 times as heavy is allotted 100. There is no upper or lower limit to the scale apart from practical limitations. As the experiment proceeds, subjects can respond to each stimulus by assigning it, if they wish, a new value. Although subjects usually find the procedure strange at first, results for both physical and attitude stimuli have proved reliable. Even a chimpanzee has used the technique successfully, in Japan (Murofeshi 1997).

In linguistics, ME has been used in vowel roughness and other speech perception experiments, where it has produced comparable but rather more useful results than ordinal scales (Fucci, Ellis and Petrosino 1990, Toner and Emmanuel 1989). Fucci et al (1990) suggest that ME could also be used by untrained listeners to measure the clarity of family members with speech communication difficulties.

In SLA research, the technique has been pioneered by Sorace and her associates. Bard, Robertson and Sorace (1996) propose that ME judgements are reliable enough to identify the relative differences in acceptability that until now have been expressed by unsystematic use of ?,??,*, ** etc. in linguistic analyses: "it delivers delicate and robust distinctions among linguistic categories" (63). It is a flexible tool for naive L1 and L2 speakers to build their own scales of acceptability without recourse to dichotomous scales where distinctions may be lost or to reference to labels affixed by the experimenter which may not match the subject's intuitions (e.g. Birdsong 1992). It then becomes possible to discriminate between "core" and "peripheral" applications of a linguistic principle. Their paper shows that this technique is at least as reliable and valid as other methods of measurement.

One other advantage of ME over other ranking methods may be in eliminating or reducing the "anchoring" or "assimilation effect" found in studies by Nagata (1992, 1997). Nagata (1997) compared judgements of target sentences presented to subjects following "anchor" sentences of varying grammaticality: ungrammatical target sentences were rated as worse if they followed grammatical "anchors" than if they followed ungrammatical "anchors" or were presented without anchors at all. This finding was robust across sentence types. In the present study, a modulus sentence of intermediate grammaticality was presented first, and remained on screen throughout the test. This acted as a kind of initial "anchor". It was apparent that subjects ceased to pay attention to it after the first dozen or so items. Given that subjects were instructed to make proportional judgements throughout, it is possible that they used each test item as an "anchor" for the subsequent item so that overall there is no assimilation effect.

ME has yet to be widely adopted in SLA, and some reservations should be expressed. Firstly it can be argued that while characterising degrees of brightness or speech clarity is common in everyday life ("dazzling", "too bright", "dim"), the same is not true of degrees of grammaticality: in everyday life utterances and written sentences are, for most people, grammatical or ungrammatical. Therefore, in experimental situations subjects may be expected to use the extreme values only, reserving the intermediate values for items that they are unsure about, or don't understand. Data presented in Bard et al (1996) show clearly however that naive respondents do not bunch items at the extreme ends of hierarchies, and that judgements. although frequently disagreeing with linguists' intuitions, are consistent across subjects. Moreover, comparisons across studies can be made more easily. Bard et al

(1996) explicitly renounce any claim that ME is *less* performance-bound than any other GJ technique, but that performance factors can be investigated: "a flexible response measure and statistical techniques like linear regression should help us to discover the major factors contributing to acceptability judgements" (64).

One of the purposes of the present study is to investigate the use of ME with a wider range of subjects than is usually found in SLA studies - including subjects who are older, less familiar with experimental conditions and less likely to have had formal instruction in the second language. The results of the informal experiment using pencil and paper reported in Bard et al (1996) strongly suggest that results obtained by this technique should be both reliable and informative.

Both SLA and FLA studies reveal a wide range of test procedures, including instructions to respondents, rating scales and timing of responses. There has been relatively little discussion of many of these aspects, with the notable exception of Chaudron's 1982 survey article and - on ranking scales - Sorace (1990) and Bard, Robertson and Sorace (1994). Much work therefore remains to be done, in particular to improve reliability without losing the procedural simplicity that allows the testing of a wide variety of subjects.

6.5   The Pilot Studies

Two small-scale pilot studies were carried out to try out the test materials and a rating method. They are described here, and the results of the second pilot experiment presented and discussed.

6.5.1 The First Pilot Experiment

This pilot study was conducted to check that the test materials were judgeable by both native and non-native speakers. Two native and two non-native speakers were tested informally. Results were not analysed.

6.5.2 The Second Pilot Experiment

For the second pilot study, 8 Italian and 3 native speakers of English were selected and tested. The Italians were recruited by word of mouth and were mainly teachers of Italian and university lecturers (in other disciplines). The native speakers were university students and staff, without any training in linguistics.

6.5.3 Experimental Hypotheses

Only adult learners took part in this experiment. The main purpose was to test the validity of the materials and to find out if they did elicit response differences between native speakers and adult learners of English, on four types of wh-movement constructions, both grammatical and ungrammatical. The small size of the sample and the absence of child learners would mean that only indicative findings could be expected.

Hypothesis 1: Highly proficient adult learners of English, including those almost indistinguishable from native speakers, will judge Subjacency and ECP violations differently from native speakers.

Hypothesis 2: Sentences instantiating the Italian parameter setting of the Subjacency Condition will be judged as more grammatical by Italian speakers of English than by native speakers.

Hypothesis 3: Italian subjects with later exposure to English (i.e. who arrived to live in Britain in their thirties) will both be judged less proficient, and diverge more from native speakers than younger arrivers.

6.5.5 The Subjects

Information on subjects' age, age of arrival, education and language learning experience was collated from the questionnaires. Ages ranged from early 30s to early 40s, with the largest number in the 36-40 age group (n=4), apart from one subject who was over 66. Average age of arrival was 29;8 (range 22-40 years). All had completed university education, and all had attended English classes in Italy, most at school and university. All spoke English at home, but Italian as well. 5 were born and brought up in Northern Italy, and three in the south (Sicily, Campania and Sardinia).

6.5.6 The Materials

79 sentences, including 9 grammatical fillers and 9 ungrammatical nonsense items were

devised. For the 4 Sentence Types - adjuncts islands, relative clauses, Noun complements and wh-islands - one control and a maximum of 4 violation sentences was prepared for each violation type and subtype (Subjacency, ECP - subject extraction, ECP - adjunct extraction). Two randomised orders were created.

6.5.7 Test procedure

Subjects were invited to an individual testing and interview session at either Strathclyde or Edinburgh Universities. Sessions lasted for about one hour and consisted of a questionnaire, to gather biographical information, the grammaticality judgement test and a conversation on a topical subject which was recorded on audio cassette, for later transcription and assessment by independent judges. The grammaticality judgement test was presented in the form of a booklet, with one item per page, followed by a box. Subjects were asked to read each sentence and to give a mark out of 6 in the box, according to how acceptable they considered it (1=possible in English, 6 = impossible in English). Although the test was not paced or timed, subjects were asked to spend as little time as possible on reading and judging the sentences, and not to look back to previous pages.

6.5.8 English Proficiency

Five minute conversations were recorded during the testing session. These were transcribed and assessed by two independent judges. The assessment procedure was adapted from White and Genesee (1996). Judges were asked to grade each subject separately on morphology, grammar, vocabulary, discourse structure and finally on "nativeness" - an overall estimate - by marking a 9 cm line labelled "beginner" at the left-hand end and "native speaker" at the right hand end. The ratings for each language level for each subject were transformed into numerical scores by dividing the lines into centimetre intervals. Scores for all language features from the two judges were added together. On the basis of these global scores subjects were divided into two groups - near-native (scores from 85/90 to 73/90) and non-native (71/90 - 56/90). The average age of arrival of the "near-natives" was 31;5 and of the "non-natives" 27;8.

6.5.9 Results and Discussion

As the number of subjects was so small, it was decided to adopt the procedure of Coppieters (1987), replicated by Birdsong (1992): a native "norm" was calculated from the scores of the three native speakers, then total divergence from the norm calculated for each subject, for

each item[4].

Table 6.1   Cumulative Divergence from NS 'Norm'

Subject No.	Age of Arrival	Proficiency Rating (90= native speaker level)	Divergence
4	40	73	-69.17
3	25	61	-64.67
7	28	75	-36.67
6	22	71	-36.67
2	30	85	-29.67
1	27	74	-26.67
5	33	56	-25.67
9	ns	90	-23.67
8	33	71	-8.67
11	ns	79	.33
10	ns	88	23.33

ns= native speaker

What is immediately obvious is that neither age of arrival nor assessed proficiency relates to cumulative divergence, but that there are no great differences - and some overlap - between native and non-native speakers, apart from subjects 4 and 3. Birdsong (1992) also found overlapping cumulative divergence scores as between native and non-native subjects (which may in fact be an artifact of this method; alternatively a function of the small sample sizer. Examination of subjects' means for types of violation shows that for most types there were no differences between native speakers' and non-native speakers' scores, but that for overall Subjacency violations, there were significant differences (F=8.07; p 0.02). Within the 4 Sentence types, statistically significant differences between these two groups were also found for grammatical extractions from relative clauses (F= 5.51; p 0.04). A main effect for proficiency (F= 5.72, p 0.03) was found for the Italian parameter setting, but numbers were too small for this to be more than indicative.

Proficiency ratings showed a wide spread, with only one non-native close to native speaker figures. However, there was no strong correlation between age of arrival and proficiency, the oldest arriver scoring in the upper half of the range. The low rating for Subject no 11, a native speaker, was due to assessment by one of the two judges, who rated him only 6/9 for fluency, and 7/9 for overall nativeness. This suggests a difficulty with the judging procedure. Using transcripts avoids judgements of syntax and vocabulary being contaminated by hearing a foreign accent, but also highlights discourse features such as hesitation, uncompleted sentences, overuse of fillers that may have been induced by the presence of the tape recorder,

or which may be a characteristic of that person's speech. Moreover, teachers of English become highly skilled at assessing adult students from their spoken production, and may find assessing transcripts less easy. Comments made by the judges seem to suggest this: "I had to keep in mind it was speech and not written work and not be too critical"; "Sometimes easy [to use to line to make judgements] because a good, connected flow but ... [Subject no 1] difficult because it improved half-way through."

To conclude, Hypothesis 1 is partially confirmed by this small-scale experiment: there are clear differences in responses to Subjacency violations by non-native speakers compared to native speakers, with general agreement on both grammatical extractions and ECP violations, which are stronger than Subjacency violations. Hypothesis 2 was not confirmed: native speakers rated "Italian" values as better than non-natives, who were also more erratic (NS mean: 2 ; SD 0; NNS mean:3.031 SD:1.6). Hypothesis 3 was also not confirmed. While only one of the L2 speakers was rated as within the range of proficiency of the NSs, there was no correlation between proficiency and age of arrival. It was decided nonetheless that the Pilot Study results were not conclusive enough to warrant a change in the focus of the research hypotheses being made for the Main Study. Age of arrival groups were too small to derive trends from grouped results. The Pilot Study results also made it clear that, for the main study, child arrivers would have to be included to ensure that no false analogy between L2 and adult speaker was created: that is, there is no consensus that child arrivers have the same intuitions as native speakers simply because they usually achieve native-like proficiency (Mack 1984, Johnson and Newport 1989,1991; Hyltenstam 1992), therefore it is important to include them in any study of age effects and L2 knowledge.

For the main study, it was decided to revise the test materials to improve their "naturalness" and to match the lexis of controls and violations more closely. We also decided to use magnitude estimation to elicit interval scales of judgements rather than asking subjects to rank items according to an ordinal scale. It was evident from the subjects' use of the ranking method in the pilot study that most subjects were willing and able to deploy all the points on the ordinal scale. Therefore, there seemed no argument against giving subjects the opportunity to make use of the greater flexibility offered by ME.

*

[1] In fact it could be argued that length is confounded with complexity in this study, as no long sentences were included which were not also syntactically complex (eg no "flat" sentences with conjoined clauses)

[2] In the case of the parameterised Subjacency violation items, this was not achieved, due to experimenter error.

[3] For this Levelt (1972) recommends minimal pairs to determine, for example, which of two rules contributes more to grammaticality: 'Thus, *a ranking of rules can be obtained by a ranking of differences in grammaticality*' (Levelt's italics)

[4] Birdsong (1992) criticised this method of calculating cumulative deviances as it does not take account of variance: "to use a mean as a norm in ... cases of great variability is to pervert the notion of norm: deviation from the mean is the norm" (fn 723). However, in most cases, there was little variability in each subject's mean scores.

CHAPTER 7

THE MAIN STUDY

7.1   Introduction

This chapter describes the experiment which was devised to test the research hypotheses. The purpose of this experiment was three-fold: (1) to discover to what extent an adult second language speaker's age on arrival in the second language community is a factor in asymptotic intuitions about wh-movement in the L2; (2) to discover to what extent native and near-native speakers of English have the same intuitions about a hierarchy of acceptability for wh-movement, as measured by the technique of magnitude estimation; and (3) to discover to what extent these intuitions match the hierarchy of grammaticality which derives from linguistic theory.

This chapter will give an account of the experiment conducted in 1997-98. It will describe the subjects who were recruited for the experiment; the test materials and procedure adopted; and present the results.

7.2 Research Hypotheses

The two sets of research hypotheses presented in Chapter 5 are repeated here for convenience.

7.2.1 Age of Arrival Hypotheses

a. In a Grammaticality Judgement test of Subjacency and the ECP scores of Italian speakers of English who arrived before the age of 15 will not differ significantly from those of native speakers, but will differ from those of adult arrivers.

b. Age of arrival in adulthood will correlate negatively with scores on grammatical and ungrammatical wh-movement sentences, such that the later the age of arrival the greater the divergence from native speaker scores.

7.2.2 Acceptability hierarchy hypotheses

a. Native speaker subjects will show sensitivity to degrees of ungrammaticality according to the "Barriers" account of Chomsky (1986). ECP violations will be rated lower than Subjacency violations, and some type of the latter will be rated lower than others, depending on the number of barriers crossed. The order of acceptability will be as follows:

Table 7.2.1   Acceptability Hierarchy for Wh-movement Constraint Violations

Violation Type	Number of Barriers Crossed	Sentence Type	Extraction Type	Identity Number	Acceptability level
Subjacency parameterized -	1	wh-island	object from finite clause	1	most acceptable
Subjacency parameterized -		wh-island	object from embedded nonfinite clause	2	
Subjacency invariant -		wh-island	object	3	
		noun complement	object	4	
			adjunct	5	
	2	relative clause	object	6	
		adjunct clause	object	7	
		wh-island	subject	8	
ECP		wh-island	subject	9	
		noun complement	subject	10	
		relative clause	subject	11	
		wh-island	adjunct	12	least acceptable
		noun complement	adjunct	13	
		adjunct clause	adjunct	14	

b. Italian speakers of English who arrived before the age of 15 will obey the same hierarchy as native speakers.

c. Adults will observe the hierarchy as far as strong violations and ECP violations are concerned, but the later the age of arrival in adulthood, the more indeterminate the judgements of weaker Subjacency violations.

d. Adult arrivers will tend to rate Subjacency violations which are grammatical in Italian but not in English higher than native speakers or child arrivers, and ratings of these violation types will correlate negatively with age of arrival.

7.2.3 Additional hypotheses

a. The grammar of an L2 speaker of English who has been resident continuously in the L2 environment for a minimum of 5 years can be deemed to have completed its development and be described as an "end-state grammar". Length of residence of these subjects will therefore not predict test performance.

b Although not all such speakers will have achieved native levels of performance proficiency in grammar, phonology or lexis, they can still have nativelike intuitions about acceptable and unacceptable sentences in the target language. Measures of proficiency above a certain level will therefore not necessarily correlate with test performance.

c. As type of input may have a bearing on the final state of the L2 grammar, subjects who

had formal instruction in English, and who use English as their main language at home and at work will perform closer to native speaker norms than those who acquired English informally and whose main language is Italian.

d. As age has been shown to correlate negatively with processing ability, subjects over 50 are likely to rate complex grammatical sentences lower than younger subjects. Subjects over 50 may also show more flexibility in responses to Italian settings of the Subjacency Condition, by a greater tendency to accept them.

7.3. The Subjects

48 subjects of Italian birth, whose first language was Italian, and who had lived in Britain for at least 5 years, were tested. 6 subjects were subsequently dropped from the study: 4 were excluded because they were found to have failed to follow the instructions for the experimental task correctly. Later, in order to produce evenly-sized age groups, two subjects were identified from scatterplots of results as giving some anomalous ratings, and were dropped [1]. Of the 42 who remained 28 were women and 14 men. The age-range was 20 -73. The median age group at the time of the test was 36-40. The average age of arrival in Britain or in an English-speaking country (one subject lived in Canada for 3 years immediately before moving to Britain) was 21 years 4 months. Most subjects were tested in individual sessions, on three sites (classrooms in Strathclyde, Stirling and Edinburgh Universities).

7.3.1 Recruitment

Subjects were recruited in a variety of ways. Useful contacts were made with the Italian Consulate General in Edinburgh, the Italian Vice-Consulate in Glasgow, the Italian Cultural Centre in Edinburgh, various Italian associations in both cities, such as Le Donne Italiane (Glasgow), and other official and unofficial bodies. Italian Departments at the Universities of Edinburgh, Glasgow and Strathclyde were also pressed into service, as were teachers of Italian in secondary schools and in community education classes in the two cities. Some Italian shops, cafes and restaurants with strong links with the local Italian community were visited.

Most subjects were contacted by word of mouth. Advertisements in university departments, appeals to churches and associations, "cold calls", letters to local newspapers and university

staff bulletins, were usually unproductive, whereas asking each respondent to suggest other likely candidates, then telephoning them would generally result in successful contacts being made.

A large number of respondents were contacted through their continuing links with Italy. 5 respondents were full- or part-time teachers of Italian, in schools, evening classes, colleges and universities. Other migrants with a strong Italian connection in Scotland were workers in the Consulate and in the Cultural Institute, translators or Italian-speaking tour guides (6). Both these groups were educated to university level, and most arrived in Scotland on completion of their university education, in their early 20s. Although they may have moved here to improve their English, their present employment requires the use of standard Italian. Apart from these, three of the subjects, two of them former "war-brides", were contacted through the Italian expatriate associations.

The largest category however comprises those who came to Scotland not to promote their language or culture but in order to take up a professional post or to study at a university. 7 subjects were undergraduates or postgraduates at the time of the study; 12 were lecturers or in other professions (photography, film, local government) (10). These individuals tend not to speak Italian in the course of their work and were contacted via friends and colleagues, rarely being linked with the migrant associations.

The final category includes those who migrated for work, often unskilled. These people usually came here as teenagers and work in Italian cafes and restaurants and shops. Some arrived as children and still work in the family business. Their schooling (in Italy) was completed by the age of 18, sometimes much earlier, and they generally learned English informally, in Britain. Although it was easy enough to call in to cafes and restaurants, they tended to be difficult to recruit for the test as they work long, hard hours.

Italians who moved to Scotland whether as children or as adults and have been absorbed into the local community, marrying British people and no longer using Italian in everyday life, proved the most elusive group (5). Three were the offspring of British-Italian marriages, who had been born in Italy, had been brought up speaking Italian, then moved to Britain when the marriages ended.

While slightly more respondents came from Southern than from northern regions, nearly every region is represented: the largest number from Tuscany, Emilia-Romagna and Lazio. This distribution matches the figures given by the Italian Consulate General in Edinburgh for the regions of origin of Italian passport holders in 1983 (the latest available figures) which show Lazio and Tuscany accounting for around 50% of the total for both Glasgow and Edinburgh (Dutto, 1986:39).

Migration dates of respondents range from the 1940s to the 1990s, and no pattern of reasons for migration emerged: marriage, study and work were reasons given, but even such broad categories are misleading. Some respondents came to marry British spouses, others came to join Italian spouses who had migrated earlier. Respondents who gave study as a primary reason may have come to improve their English, to go to boarding school or to take up a postgraduate degree course. "Work" encompassed compulsory farmwork (an imposition on all unskilled migrants until the late 1950s), restaurant waiting and university lectureships.

7.3.2 Biographical Information

a. Age of Arrival   Factors visible in reasons for migration to Britain are reflected in the pattern of respondents' ages on arrival. Ages of arrival cluster around what could be described as significant life stages: leaving school (8 arrived between the ages of 17 and a half and 19); graduation (7 arrived aged 22 and 23). There is a third cluster around 25 and 26 (7 respondents) - possibly later graduation or a career move. Another 7 moved between the ages of 29 and 31, giving as reasons work, marriage and family circumstances. It should be noted that the subjects in the present study all had a single age of arrival, unlike those in Birdsong's 1992 study who tended to move between France and the USA at different stages in their lives. Our subjects returned to Italy frequently, but only for holidays.

b. Age at Time of Test.   At the testing session, subjects were asked to tick the age-group they belonged to (20-24, 25-29, etc), so a precise analysis of subjects' age is not possible. There were more subjects in their 30s than in any other decade (13) and only 9 aged 50 to 73.

c. Length of Residence.   Respondents were selected on the basis that they had lived in the UK, or in an English-speaking country for a minimum of 5 years. However two were included in the study who had lived here for 4 years, as their spoken English was judged to

be of near-native standard. 17 had lived in Britain for less than 10 years; 11 for between 11 and 25 years and 14 for more than 26 years.

d. Education. 27 respondents were educated to university level, and several had higher degrees, or had continued studying well into their 20s (even 30s). 7 left school before 18, and 8 completed high school. Although it would have been preferable to have tested only people who had at least completed high school, the difficulty of finding respondents, and of finding enough respondents for each Age of Arrival group made such a stipulation impossible to satisfy.

e. Proficiency in English. Having been told of the criteria, which included the ability to "speak English as well as someone who was born here" and literacy in English, some of those approached declined, on the grounds that their English was not adequate. 29 of the respondents were assessed by 4 independent judges from telephone conversations, or taped conversations. The assessment procedure was adapted from White and Genesee (1996). Two judges assessed each subject. Each judge completed a form for each subject commenting on his or her grammar, vocabulary, pronunciation and fluency, then marked a point in a 9 cm line labelled at either end "beginner" and "native speaker". Judges were experienced English Language teachers. From the 29 who were assessed (all adult arrivers), 6 (20.7%) were judged to be below the mid-point of 4.5cm; 13 (44.8%) were marked between 5cm and 8.5cm; 10 (34.5%) were marked above 8.5cm.

The remaining 13 subjects, who for a variety of reasons could not be assessed in the way described above, were judged by the experimenter on the basis of telephone and face to face conversations before and during the testing sessions. Of this group, 6 were child arrivers whose spoken production was indistinguishable from native speakers.

It was decided to classify subjects into two categories: near-native and non-native. The impressionistic character of the assessment method discouraged further subdivision although it might be argued that the wide spread of marks awarded by the judges requires three groups. As the main purpose of the assessment was to allow controlling for proficiency, separating the near-natives from the non-natives was considered sufficient. In the former category (n=20) were placed those marked above 8.5 cm by the independent judges, together

with the 6 non-assessed child arrivers and 4 of the non-assessed adult arrivers. These were individuals whose spoken production was considered equal to that of a native speaker. Accent was not taken into consideration, although stress and intonation was. The remaining subjects (n=21), placed below 8.5cm on the line, were assigned to the "non-native" category.

f. How English was acquired. Respondents were asked how they had acquired their skill in English. Most had attended classes at some stage: 18 learned English at secondary school, for periods ranging from 4 to 7 years; 16 studied it at university or at a private language school; and 15 attended classes for short periods in the UK or USA (5 *only* attended classes in the UK). 12 was the earliest age at which formal instruction began. 15 subjects had no formal instruction in English, including all 8 child arrivers. It was decided not to compare ages of first exposure as many of the older subjects who had had formal instruction found it difficult to remember exactly when they first started English classes, or to quantify this. Several studies (for example Newport and Johnson 1989, Birdsong and Molis 1997) have found no correlation between test performance and age of first exposure.

g.Use of English and Italian. Two questions in the questionnaire asked how respondents used their two languages. At home, 18 respondents spoke mainly English and 18 spoke both languages; only 6 spoke mainly Italian. 17 subjects used Italian in formal work domains such as teaching, tour-guiding or translating.

7.3.3 Native speaker controls

9 native speakers acted as controls in the experiment. They were recruited from staff and students at Strathclyde University. 8 were educated to university level, one to secondary school level, and one had a higher vocational qualification. None had studied linguistics. Ages ranged from 22 to 51, the modal age group being 26-30.

7.3.4 Age of Arrival

5 Age of Arrival groups were formed. The cut-off points for each group were fixed to correspond to maturational stages and significant life events. It was also considered important to form groups of roughly equal size, to improve the validity of the statistical analyses. Group 1 comprises individuals who arrived in Britain as children, that is, from 3, after the start of the critical period and when the first language may be considered to be

established[2] up to 14. Other studies have chosen 16 or even 17 as the cut-off age for child arrivers, but 14 is closer to the onset of puberty, and was in fact the age Lenneberg (1967) identified for the end of the critical period. It is also within the age-range for compulsory schooling in Italy and Britain, as far as the older subjects are concerned.

Group 2 includes those who arrived after completing their schooling - the majority around 18 years of age. Some came to work or to marry, others to enter or prepare for higher education. They all came without their families, and can be considered young adults. Most of the subjects in Group 3 arrived around the age of 22, often after graduating and specifically to improve their English. Later arrivals do not fall conveniently into "demographic" categories, but it was considered useful to continue with the five year age groupings.

Table 7.3.1   Age of Arrival Groupings

Group Number	Age Range	Number of Subjects	Average Age of Arrival
1	3-14	8	7;4
2	15-19	8	18
3	20-24	9	22;4
4	25-29	9	26;6
5	30-34	8	31;9
6	native speakers	9	-

7.4.5 Biographical Factors

Seven biographical factors were derived from the information provided by the non-native subjects in the pre-test questionnaires, as summarized above. These factors were operationalised for use with statistical procedures as follows:

(1) CURRAGE (Age at time of test in years)
  1   20-29 (10)
  2   30-39 (13)
  3   40-49 (10)
  4   50 + (9)
(2) LOR (Length of residence in years)
  1   5-10 (17)
  2 11-15 (6)
  3 16-20 (2)
  4 21-25 (3)
  5 26+ (14)
(3) EDUC (Age education completed)
  1 17 (7)
  2 18-20 (8)

3 21+ (27)
(4) PROF (Proficiency in English)
 1 near-native(20)
 2 non-native (22)
(5) MODE (The way English was acquired)
 1 Informal (17)
 2 Formal (25)
(6) HOME (The main language spoken at home)
 1 English (18)
 2 Italian (6)
 3 Both (18)
 (7) ITUSE (Whether Italian is spoken at work, in formal contexts such as teaching)
 1 Yes (17)
 2 No (25)

7.4 Test Materials

The test consisted of 82 sentences, presented in two randomized orders (see Appendix II). As outlined in Chapter 4, four types of wh-movement construction were selected:

Type 1: Extraction from Adjunct Clause

Type 2: Extraction from Relative Clause

Type 3: Extraction from Noun Complement

Type 4: Extraction from Wh-island

Each Sentence Type had grammatical control items and varying numbers of items instantiating types of Subjacency and ECP violations. These are named Extraction Types. The Extraction Types selected for each Sentence Type are shown in Table 7.4.1:

Table 7.4.1:   Test Materials

Sentence Types	Extraction Types						
	grammatical	parameterized subjacency	invariant subjacency 1 barrier - obj	invariant subjacency - 2 barriers - obj	invariant subjacency- adj and subj	ecp - subject	ecp - adjunct
1: Adjunct Clause	x			x			x
2:Relative Clause	x			x		x	
3:Noun Complement	x		x		x (adj)	x	x
4 Wh-island	x	x	x		x (subj)	x	x

4 different lexicalizations were devised for each Sentence Type, so that grammatical controls matched violations, as in (1) (Adjunct Clause)

(1a) Didn't Ann want to go home after drinking a bottle of vodka with her friends?

(grammatical control)

(1b) What did Ann want to go home after drinking a bottle of with her friends (Subjacency violation - extraction from adjunct)

(1c) After doing what with her friends do you wonder whether Ann wanted to go home? (ECP violation - extraction of adjunct)

(1d) Where did Ann want to go home after drinking a bottle of vodka with her friends (ECP violation - extraction of adjunct from adjunct)

7.5 Test Procedure

16 subjects were tested in Edinburgh University, 24 in Strathclyde University, and 2 in Stirling University. Native speakers were tested at Strathclyde University. Sessions lasted for approximately 45 minutes. 7 of the native speaker controls were tested in a group, but most of the Italian speakers were tested individually, as they became available. No sessions involved more than two subjects. Each session began with a brief explanation by the experimenter of the nature and purpose of the experiment, based on the script (see Appendix I.1). Subjects then completed the questionnaire (Appendix I.2) on their language background and carried out the two practice tests, after reading the instructions for each. The experimenter used this part of the session to converse informally with subjects in order to assess their oral proficiency. Before each practice the instructions were rehearsed orally, to check understanding. After each practice each subject was asked how s/he felt about using the method of magnitude estimation.

The first practice involved subjects giving numerical estimations of line length. Subjects were shown 10 randomly presented lines, ranging from 20cm to 1cm in length, on an OHP screen. The modulus line was 10cm in length, and remained on screen throughout. Subjects were told that the modulus line was in the middle of the range. Subjects wrote their estimations in the test booklet. In the second practice, 6 short sentences were displayed on the screen, the sentences instantiated fairly basic word order and grammatical violations. Subjects were asked to give numerical estimations, this time of grammaticality or acceptability, awarding higher numbers to more grammatical items. The modulus sentence - mildly unacceptable - remained on screen throughout, and subjects were told that this sentence was in the middle of the range.

Immediately following the practice tasks, subjects were asked to read the instructions for the main test (Appendix I.1), which were almost identical to those for the second practice. The

same information regarding the modulus sentence was relayed. Subjects were informed that they would have enough time to read each sentence, but should try not to work out missing words or make corrections. The test was not timed, but each new sentence was presented on the OHP screen as soon as the subject had written a number for the previous sentence. All subjects were offered a few minutes' rest after item 35, and half did take a break.

A few subjects commented at the end on their performance ("I don't know if I've done it right") but none expressed unhappiness with the method of magnitude estimation.

## 7.6   How Magnitude Estimation Was Used

In order to show that subjects were able to use this unfamiliar technique for judging, data from the practice tasks are presented here. In the first practice subjects were clearly told that their mathematical skills were not being tested, and that they were to give numbers to lines based on their impression of each line length as it related to the first line. They were also encouraged not to restrict themselves to conventional scales such as 1-10 or 1-100. Subjects used a wide variety of scales: only two limited themselves to 1-10. Figure 7.6.1 is a scatterplot showing logged estimates plotted against logs of actual line lengths: it can be seen that subjects were relatively accurate in their estimates. The Pearson product-moment correlation (r) measures their association according to the assumption that the function relating the psychological axis to the physical axis is a straight line. $r = 0.9902$ (maximum value is 1.00).

Figure 7.6.1: Scatterplot of numerical estimates of line lengths
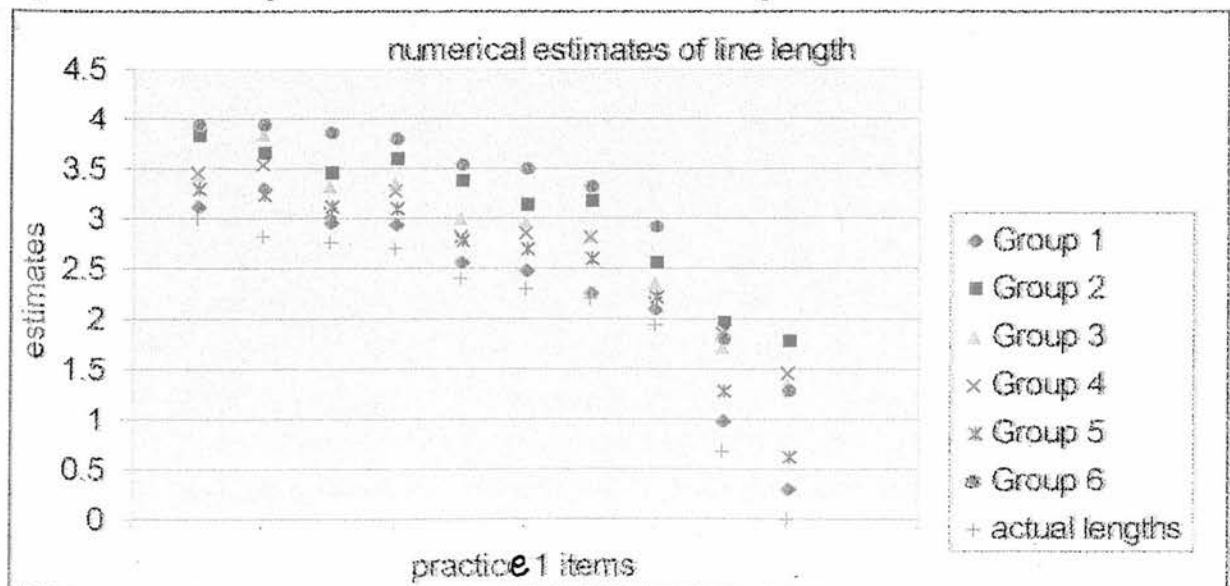
The second practice - numerical estimations of grammaticality with 5 short sentences - produced a less clear picture. Here, subjects were agreed on the extremes, which were a

Figure 7.6.2: NS and NNS scores of practice 2 items
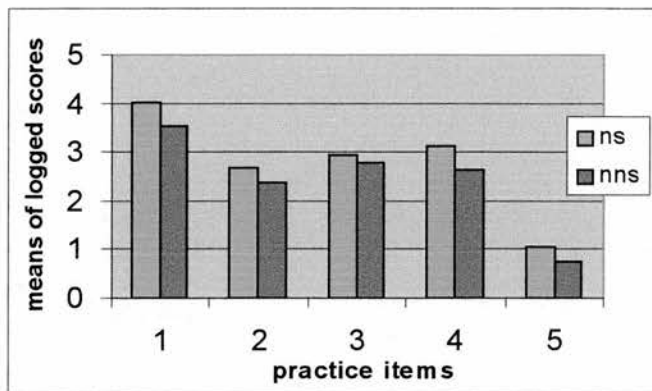
nonsense sentence and a grammatical sentence, but they were not consistent in ranking the intermediate sentences, which instantiated overt subject, word order and agreement violations. There were no differences between native and non-native speakers (r=.9904)

In the main task, subjects mostly did not limit themselves to conventional 10-point scales - that is, they followed the instructions. Even those without experience of higher education were able to use the method successfully. The 4 subjects who were dropped at this stage were found to be either scoring randomly throughout, or to have changed halfway through the task from a a wide of range of numbers to only two (100 and 1; 10 and .1) for all items, ie to a nominal scale. Consistency of scoring can be assessed, rather approximately, from the variance in the means for grammatical and nonsense filler items: the overall means for grammatical fillers was 3.82 (SD 1.34) and for nonsense fillers 0.99 (SD 1.84).

7.7   Results

7.7.1 Descriptive statistics

Raw scores were logged in order to produce comparable scales for all subjects. Means were then calculated for extraction types within the four Sentence Types. Group means for all sentence types are shown below.

Table 7.7.1   All Grammatical Items

AOA Group	3-14	15-19	20-24	25-29	30-34	NS
GRAMMATICAL ITEMS	4.01	4.12	3.57	3.70	3.37	3.65
UNGRAMMATICAL ITEMS	2.52	3.21	2.59	2.80	2.40	2.24

Table 7.7.1 shows mean scores of the five Age of Arrival groups and native speakers on all grammatical and all ungrammatical test items. A paired-sample t-test gives significant results by grammaticality for all groups (t=10.204, df=5, p <0.005), showing tht all groups differentiated between grammatical sentences and violations.

Table 7.7.2 Sentence Type 1: Adjunct Extractions

EXTRACTION TYPE	AoA Group						
	3-14	15-19	20-24	25-29	30-34	NS	ALL
GRAMMATICAL CONTROL	3.897	4.120	3.612	3.738	3.440	3.714	3.750
SUBJACENCY VIOLATION - object	3.897	3.467	2.639	3.060	2.678	2.619	2.862
ECP VIOLATION - adjunct	2.552	3.380	2.658	2.804	2.408	2.375	2.691

All Age of Arrival groups rated controls as considerably better than violations, and all but the 20-24 group rated Subjacency violations slightly better than ECP violations.

Table 7.7.3 Sentence Type 2: Relative Clause Extractions

EXTRACTION TYPE	AoA Group						
	3-14	15-19	20-24	25-29	30-34	NS	ALL
GRAMMATICAL CONTROL	3.917	4.166	3.554	3.597	3.275	3.509	3.663
SUBJACENCY VIOLATION - object	2.392	3.097	2.379	2.777	2.202	1.776	2.430
ECP VIOLATION - subject	2.459	2.877	1.442	2.606	2.231	1.171	2.390

As shown in Table 7.7.3, all groups rated violations considerably worse than grammatical controls. Subjacency violations were judged to be better than ECP violations by native speakers and by the younger adult groups.

Table 7.7.4   Sentence Type 3: Noun Complement Extractions

EXTRACTION TYPE	AoA Group						
	3-14	15-19	20-24	25-29	30-34	NS	ALL
GRAMMATICAL CONTROL	4.116	4.270	3.654	3.749	3.489	3.713	3.824
SUBJACENCY VIOLATION - object	2.730	3.305	2.467	2.848	2.158	2.149	2.602
SUBJACENCY VIOLATION - adjunct	2.275	3.089	2.396	2.520	2.330	2.166	2.457
ECP VIOLATION - subject	2.846	3.144	2.542	2.790	2.515	2.432	2.704
ECP VIOLATION - adjunct	2.208	3.258	2.740	2.813	2.672	2.361	2.673

All violations were judged to be worse than the grammatical controls by all groups. All groups also rated ECP violations to be better or no worse than Subjacency violations.

Table 7.7.5 Sentence Type 4: Wh-island Extractions

EXTRACTION TYPE	AoA	Group					
	3-14	15-19	20-24	25-29	30-34	NS	ALL
GRAMMATICAL CONTROL	4.097	3.928	3.477	3.702	3.29	3.676	3.690
SUBJACENCY VIOLATION - parameterized, finite clause	3.433	3.906	3.214	3.285	2.916	3.068	3.297
SUBJACENCY VIOLATION - parameterized, nonfinite clause	2.782	3.502	2.855	3.118	2.687	2.313	2.869
SUBJACENCY VIOLATION - object	2.095	3.200	2.622	2.716	2.136	2.381	2.528
SUBJACENCY VIOLATION - subject	2.022	2.872	2.104	2.490	2.032	1.555	2.171
ECP VIOLATION - subject	2.028	2.645	2.931	2.481	1.972	1.719	2.301
ECP VIOLATION - adjunct	2.536	3.119	2.725	2.899	2.554	2.519	2.725

All groups rated invariant Subjacency violations as worse than grammatical controls, and extractions from and of subjects worse than extractions of objects. Extractions of adjuncts were judged to be much better than even object extractions by all subjects, most notably native speakers and child arrivers. The parameterized violations were rated lower than the controls but higher than the invariant settings, with the non-finite value considered to be less acceptable than the finite value. Only the native speakers rated the nonfinite value as low as invariant Subjacency violations.

Age of Arrival group means for all sentence types show a clear picture as regards controls and violations: all groups judged grammatical controls as better than violations. However, as between Subjacency and ECP violations, where predictions from linguistic theory and findings from previous research would lead one to expect respondents to prefer object extractions to subject or adjunct extraction, the data presents a murkier picture. Adjunct extractions in particular seem to have been more acceptable than either object or subject extractons, *contra* "Barriers".

7.7.2 Analyses of variance

MANOVA tests were carried out for all Sentence Types. The factors tested for were Age of

Arrival and Extraction-type, and interactions between these factors.

Table 7.7.6 Age of Arrival (between subjects) effects

	F value	significance
Adjunct Extractions	.37	.866 (ns)
Relative Clause Extractions	.46	.802 (ns)
Noun Complement Extractions	.35	.878 (ns)
Wh-island Extractions	.40	.844 (ns)

Table 7.7.6 shows that for no Extraction type were there any significant differences between mean test results of each Age of Arrival group (including native speakers).

Table 7.7.7 Extraction Type (Within subjects) effects

	F value	significance
Adjunct Extractions	95.75	.0001***
Relative Clause Extractions	78.64	.0001***
Noun Complement Extractions	63.45	.0001***
Wh-island Extractions	53.08	.0001***

Table 7.7.7 shows highly significant effects of extraction for all Extraction Types. Thus all subjects clearly recognised a difference between grammatical and ungrammatical wh-movement, for both weak and strong violations. It may be noted that the F values for Sentence Types 1 and 2 (involving strong violations) are slightly greater than those for the weaker types.

Table 7.7.8 Age of Arrival x Extraction Type effects

	F	significance
Adjunct Extractions	.97	.477(ns)
Relative Clause Extractions	.90	.533 (ns)
Noun Complement Extractions	1.17	.281 (ns)
Wh-island Extractions	1.57	.034*

= significant at .05 level

Table 7.7.8 shows a significant interaction effect of Age of Arrival x Extraction Type for Type 4, Wh-islands, which proves to be due to scores for the finite clause parameterized or "Italian" version of Subjacency. Post-hoc analysis reveals that differences on this sentence type between child arrivers and the 15-19 group (although not between child arrivers and older Age of Arrival groups) are significant (post-hoc Tukey HSD test $p = <0.05$). This result is predicted by the research hypothesis; although there are no significant differences between Age of Arrival groups or between Italian and British speakers of English on ECP and strong Subjacency violations, adult learners judge this weakest type of Subjacency violation differently from child learners and native speakers. An examination of the data shows that 17 out of 34 adult arrivers and 2 out of 8 child arrivers (but no native speakers) rated this "Italian" value as high as or higher than control sentences. For extractions from nonfinite clauses, the figures were identical, although a slightly different set of test subjects was involved. In the latter case, the overall gap between ratings for grammatical controls and ratings for parameterized violations was wider (see Table 7.7.10).

7.7.3 Strength of Preference Scores

As the purpose of magnitude estimation is to measure subjects' *relative* judgements of different types of violation by asking them to set up their own interval scales, the experimenter can compare not only the means of different groups but also the width of the "gap" each group establishes between extraction types within each Sentence Type, given that items are matched lexically, as here (Bard et al 1996:50). When scores for Subjacency violations are subtracted from scores for grammatical controls, it can be seen (Table 7.7.9) that, for Extraction Types 2, 3 and 4, the child arrivers behave most like the native speakers, in that these two groups have stronger preferences for grammatical extractions than the adult arrivers. The weaker preferences of the latter suggest more indeterminate judgements.

Table 7.7.9 Strength of Preference for Grammatical Controls over Invariant Subjacency Violations

EXTRACTION TYPE	AoA Group						
	3-14	15-19	20-24	25-29	30-34	NS	ALL
Adjunct Extractions	1.1514	0.6533	0.9737	0.6786	0.7622	1.0948	0.8874
RelativeClause Extractions	1.5255	1.0690	1.1768	0.8198	1.0730	1.7330	1.2334
Noun Complement Extractions	1.3858	0.9637	1.1866	0.9005	1.3313	1.5643	1.2217
Wh-island Extractions	1.7192.	0.7282	0.8550	0.9857	1.1532	1.2951	1.1626

MANOVAs run on these scores give an Age of Arrival effect on all Strength of Preference scores close to significance (Between-Subjects F = 1.90 p 0.113) and an effect of Strength of Preference by Sentence Type slightly less close (Within-Subjects F =1.26 p 0.194).

Strength of preference scores show a similar pattern across all Sentence Types: child arrivers and native speakers have consistently higher scores (stronger preferences) compared to the adult groups, although this is not statistically significant. As differences between adult arrivers' groups were not only not significant, but not ranked according to age, it was decided to regroup subjects into two larger categories: "native-speaker+child arrivers" (n=17) and "adult arrivers" (n=34) and compare their mean strength of preference scores. MANOVAs produced significant F values for Sentence Types 1, 2 and 4 (Type 1: F=4.63, p .036; type 2: F=7.76, p .008; type 4: F=7.65, p .008).

Strength of preference scores were also separately calculated for the parameterized wh-island extractions (the "Italian" values).

Table 7.7.10 Strength of Preference for Grammatical Controls over Parameterized Values

EXTRACTION TYPE	AoA	Group					
	3-14	15-19	20-24	25-29	30-34	NS	ALL
Finite clause	0.6641	0.0222	0.2626	0.4171	0.3741	0.6078	0.3936
Nonfinite clause	1.3153	0.4270	0.6213	0.5843	0.6026	1.3630	0.8189

Group differences were significant for extractions from finite clauses (F=2.56; p 0.040), and from non-finite clauses (F=3.39 p 0.011). Post-hoc analysis shows that differences between child arrivers and 15-19 year-olds for finite clause extractions are significant (post-hoc Tukey HSD test: p = <0.05) as are those between native speakers and 15-19 year olds for non-finite clause extractions (post-hoc Tukey HSD test p=<0.05). Older Age of Arrival groups took a more indeterminate stance, neither clearly rejecting nor clearly accepting the Italian value.

7.7.4 Extraction Domains

Extraction domains were also compared. Previous studies have identified differences between subjects' responses to ungrammatical object extractions (Subjacency) compared to subject extractions Subjacency + ECP), and between extractions of objects and subjects from weak Subjacency islands (Wh-islands and Noun complements) and strong islands for Subjacency (Relative Clauses and Adjuncts). The figures presented here exclude scores from

parameterized values.

.

Table 7.7.11 Age of Arrival Groups' Mean Scores for Object and Subject Extractions

EXTRACTION TYPE	AoA	Group					
	3-14	15-19	20-24	25-29	30-34	NS	ALL
Object - weak	2.46	3.24	2.56	2.77	2.14	2.29	2.58
Object - strong	2.39	3.10	2.38	2.78	2.20	1.78	2.43
Subject	2.42	2.87	2.42	2.61	2.22	1.9	2.40

No significant differences were found between ratings for subject and object extractions across groups or within groups[3].

7.7.5   Intervening variables

Seven biographical factors were selected from the information provided by the questionnaires administered during test sessions. These were: age at which formal education was completed; current age; length of residence; proficiency in English; mode of learning English; use of Italian at work and main home language.

No significant effects on mean ratings of sentences overall were found for any of these factors. A weak effect of age at time of test was found for strength of preference in Sentence Type 1 (F= 3.18, p 0.050). Some slight, but mainly non-significant, effects of current age, mode of learning English and proficiency in English were found for some other strength of preference scores: for Relative Clause Extractions, mode of learning English was close to significance (F=2.72, *p* 0.076) showing that formal instruction correlated quite highly with larger strength of preference scores. Proficiency too impacted on this Sentence Type: F=2.65, *p* 0.081. Only for mean ratings of the finite parameterized violation was there a clear effect of main language in use (F=3.55, *p* 0.038), subjects whose main language was English tending to score them as ungrammatical.

7.7.6   Results for an Acceptability Hierarchy

We now turn to the evidence provided by the test results that subjects made judgements according to the hierarchy of grammaticality specified in the theory. Figure 7.7.1 shows native and non-native means for all extraction types. Numbers in brackets below each bar give the order predicted by the theory (see Table 7.2.1)

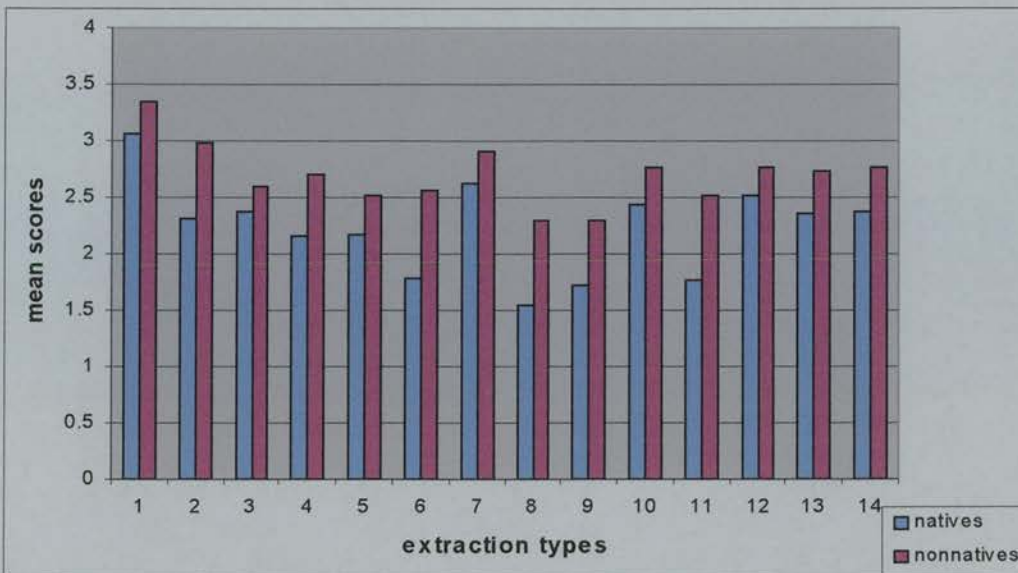Figure 7.7.1 Native and Non-native scores for all extraction types

Figure 7.7.1 shows that both native and non-native subjects rated finite parameterized Subjacency violations (no. 1) significantly higher than invariant Subjacency violations and ECP violations (nos.3-14), (by Tukey test: (>= 0.1108 for range: 4.95; *p* 0.05).

However from the overall means it is also clear that subjects do not discriminate between Subjacency violations and ECP violations. Adjunct extractions were rated unexpectedly high. This anomaly is discussed in Chapter 8

Comparing the ranking shown in Figure 7.7.1 with the hierarchy given previously (Table 7.2.1) it is clear that both native and non-native speakers are only partially observing the order proposed by the theory. There is a strong tendency to rate adjunct extractions, whether Subjacency or ECP violations, higher than object or subject extractions. Although the order is no more than indicative, as the ranking is not statistically significant, it may be noted that it is the *adjunct* extractions alone that contradict the theoretical predictions. All subjects generally prefer object to subject extractions, and weaker to stronger Extraction Types, in accord with previous studies (Martohardjono 1992, Uziel 1993).

Overall strength of preference scores for each Sentence type (for controls over Subjacency violations) (see Table 7.7.9) confirm that subjects differentiated, although not to a significant extent, between extractions from Relative Clauses, which entail strong violations, and the weaker Wh-island extractions (paired sample t-test: t=1.30, df=50, *p* 0.198), with Noun

Complement extractions lying somewhere between. The strength of preference score for adjunct islands was very low.

*

---

[1]Snow and Meijers (1977:175) point out that if subjects make errors with the most acceptable and the least acceptable test items, their judgements can be safely discarded.

[2]McLaughlin 1978 distinguishes between acquisition of two languages before the age of three, termed "simultaneous" acquisition, and "successive" acquisition when the second language is learned after that age.

[3] Strength of preference for grammatical object extractions over strong - but not weak - Subjacency violations were significant when three groups (child arrivers, younger adults, older adults) were compared (F=4.37, p 0.018). A post-hoc Tukey test shows significant differences (p 0.05) between child arrivers and older adults. Here, as means for object versis subject extraction are calculated from all Sentence Types, ie from all lexicalisations, strength of preference scores may be less convincing.

CHAPTER 8

DISCUSSION

8.1   Introduction

8.2   The Effect of Age of Arrival

8.3   An Acceptability Hierarchy for Wh-Movement Violations

8.4   Conclusion

8.1   Introduction

This chapter presents a discussion of the results of the experiment described in the last chapter. The discussion deals firstly with the results concerning the effect of age of arrival on test performance, to what extent they support the research hypotheses, possible reasons for and implications of the findings and how they compare with previous studies. Secondly, the results which relate to a hierarchy of acceptability are analysed similarly. Finally, problems, limitations and implications of the experimental results are outlined, together with possible alternative explanations for some of the findings.

8.2   The Effect of Age of Arrival

8.2.1   Summary of Key Findings

For all four Sentence Types - Adjuncts Islands, Relative Clauses, Noun Complements and Wh-islands - the experimental results show that all subjects in all age of arrival groups recognised both grammatical and ungrammatical wh-movement. There were no significant - or near-significant - differences among age of arrival groups, whether between older and younger adults or between children and adults.

In the case of English sentences instantiating the "Italian value" of the Subjacency parameter (extraction from finite clause), there was an age of arrival effect: adult arrivers were much more likely to rate these items as high as, or even higher than, the grammatical control items, although only the gap between child arrivers and young adults was significant. In addition, there was some effect of main language used at home. For other types of violation, age of arrival had no effect, either on judgements of invariant Subjacency or on those of ECP violations. All groups also made comparable judgements, unaffected by age of arrival, of weak and strong Subjacency violations and of extractions from different domains.

Comparisons of strength of preference scores for Adjunct islands, Relative Clauses and Wh-islands between child arrivers and adults (collapsed from the Age of Arrival groups) did

achieve statistical significance, as did strength of preference scores for object extractions, with child arrivers giving stronger preference to grammatical controls over Subjacency violations than adult arrivers. As intervening variables had only a slight influence on these scores, most of the difference found between subjects was due to whether they were under or over 15 when they were first exposed to English. Comparing strength of preference for grammatical control sentences over Italian settings of Subjacency across all six Age of Arrival groups also showed a main effect of age of arrival: the later the age the weaker the preference for the grammatical items.

8.2.2   Research Predictions

The research hypotheses for age of arrival predicted that subjects' test performance with respect to direction of judgements on all Sentence Types would correlate inversely with the age at which subjects arrived in Britain to live. This prediction has not been borne out. There were no significant differences in scores among the five Age of Arrival groups for invariant Subjacency or for ECP violations, although the strength of preference figures show a critical period effect - a significant difference between child and adult arrivers. It would therefore appear that, for highly proficient learners with an L1 instantiating wh-movement, the ability to distinguish between licit and illicit wh-movement items in the L2 does not depend on age of arrival in the L2 community. Once such learners realise that English has wh-movement, they also realise that the Subjacency Constraint holds, as in the L1. Uziel (1993) was also able to show that advanced Hebrew learners of English could perform like native speakers on strong Subjacency violations, as could similar groups of Italian speakers (Martohardjono 1992). Both Italian and Hebrew have wh-movement, and share most of their Subjacency features with English, so that learners of English need only transfer the L1 Subjacency parameter setting to the L2 to achieve accuracy in most instances. These results point to UG access via the L1 for adult learners.

However, adult and child arrivers did diverge on the weakest violations - parameterised object extraction from Wh-islands. On this, child arrivers behaved like native speakers - on the whole (two child arrivers did rate these violations as being as good as the grammatical controls, suggesting some L1 influence, as found by Mack (1984)). This finding supports the

hypothesis that adult arrivers will differ from child arrivers most on Subjacency violations where the L1 and L2 have different values. Strength of preference scores for finite parameterised Wh-island violations confirmed this: most subjects rated finite parameterised violations higher than invariant violations of Subjacency, but adult arrivers were more tolerant of the former, and more likely to rate these sentences higher than controls.

The strength of preference figures reveal that no groups made a significant distinction, as prescribed by the theory, between different weak and strong Subjacency violations. The strength of preference data need to be treated with caution, as has been mentioned: controls and Subjacency items were not exactly matched pairs, although they were lexically comparable, so strength of preference scores cannot be confidently read as measuring only a preference for one syntactic form over another - it was not found possible to control completely for lexis and semantic content, as explained in Chapter 6 above.

A scrutiny of the results for different Sentence Types reveals some predictable but also some surprising patterns. We predicted that Adjunct Island violations would evoke lower scores than ungrammatical extractions from Noun Complements and Wh-islands. However ungrammatical Adjunct Island extractions - which should match Relative Clause extractions for unacceptability - were rated high by all subjects.

For all Sentence Types except Noun Complements, all groups found Subjacency violations slightly more acceptable than ECP violations. However, none of these within-group differences was significant.

8.2.3 The Role of Proficiency

As stated above, the precise determining of the proficiency of near-native L2 speakers is fraught with difficulties: there is no standardised test, and studies that have attempted to assess high-level proficiency have used more or less subjective methodologies. However, those that have, like White and Genesee (1996) and Martohardjono (1992), are the ones that have found not an age effect, but a proficiency effect. In the present study, a relatively informal assessment was employed (although no less rigorous than methods used in the other studies discussed in Chapter 3) and we cannot do more than suggest the effect proficiency

may or may not have exerted. According to the informal classification made, the child arrivers' group contained only "near-native" subjects, while the 30+ group included no near-natives. Younger adults (16-19 year olds) included only one subject assessed as "non-native". Adults who arrived in their 20s (groups 3 and 4) however showed no neat gradation of proficiency according to age. It will be recalled that the second Pilot Study, with a different assessment system, produced similar findings.

In the test, however, even the 30+ group made direct judgements comparable to native speakers. The finding that proficiency had no significant effect on most violation types clearly contrasts with White and Genesee (1996) who claim a strong correlation between near-native proficiency and near-native test performance. One reason for this could be that the proficiency assessment instrument seriously underestimated some subjects' abilities. Another possibility is that the fact that English and Italian share the same settings of the Subjacency parameter for most island types makes it comparatively easy for Italian speakers to recognise Subjacency violations in English. However, this conflicts with Johnson (1988) who found that Spanish learners of English (whose L1 has the same parameter setting for Subjacency as Italian) had only "probabilistic" knowledge of Subjacency. On the other hand, White and Juffs (1998) replicated the experimental study of White and Genesee (1996) with two groups of highly proficient but not near-native Chinese-speaking adults learners of English (whose proficiency was not in fact assessed), and found little variation between them and native speakers. Moreover, one group consisted of people who had never left the People's Republic of China and had learned English almost exclusively in classroom environments; the second group had an average length of residence in Canada of 4.1 years (range 2.3 - 8). Both groups were first exposed to English in naturalistic settings after the age of 16. Rather astoundingly perhaps, especially as Chinese is not a wh-movement language and therefore does not instantiate the Subjacency Condition, neither group differed significantly from native speaker control groups on Subjacency violations; although there were such differences in the case of grammatical wh-movement items. White and Juffs propose that processing difficulties may be responsible for this latter finding, as their question formation test showed the Chinese speakers had acquired long-distance wh-movement. The fact that such speakers were able to make judgements comparable with native speakers does lend support to empirical findings (eg Snow and Meijers 1977) and

anecdotal evidence (eg Patkowski 1980) that adult L2 speakers can produce L1-like judgements without near-native mastery. Snow and Meijers (1977) found strong correlations between "excellent bilinguals", poorer speakers and native speakers of Dutch, who were asked to judge sentences with marked word order. They claim their results "strongly support the notion that syntactic intuitions in a second language are produced by a linguistic faculty which is separate from the faculty of speaking and understanding" (174).

.

8.2.4   The Role of Biographical Factors

None of the biographical factors that were controlled for in the experiment proved to have a significant impact on test scores. Thus neither length of residence (over 5 years), mode of instruction nor age at time of test affected test performance. This follows the findings of other studies. Johnson and Newport (1989,1991), with no proficiency criteria, found that length of exposure and a range of factors relating to language instruction failed to correlate significantly with test score. Birdsong and Molis (1997), who replicated the first J&N study with Spanish speakers of English state that "the best predictor of accuracy, aside from A[ge o[f] A[rrival], is English use", (r=0.446, p 0.0004) not type of instruction. English use however did influence ratings of parameterised violations: subjects who spoke English as their main language were slightly more likely to perform like native speakers on the finite parameterised violation compared to those who reported using both languages or mainly Italian. Use of English was defined in this study as "main language spoken at home" and it was not possible to measure accurately the quantity or "quality" (eg the range of domains) of that use. Closer investigation of subjects' use of their L2 as well as their proficiency in future studies may add to our understanding of L2 grammars: ultimate attainment research has so far tended to focus rather exclusively on quantifiable metalinguistic performance, and to perhaps to discount aspects of use and proficiency which, with adult subjects in particular, will have to rely on self-report and qualitative data.

Bialystok , in her critique of J&N (1997), argued that the striking degree of variance amongst adult arrivers compared to younger learners evident in their data is due to the latters' experience of US school education. This factor cannot of course be separated from age of arrival, but it is difficult to divine *which* aspect of school education could produce such

homogeneous results in metalinguistic tests. Bialystok (1997) mentions "explicit language training more characteristic of earlier grades" (123): child arrivers in the present study had no ESL training in school, and most attended school at a period when grammar instruction, and a focus on accuracy in writing was out of fashion.

Subjects' age at the time of the test also failed to interact significantly with test performance, although strength of preference figures for subjects over 50 were consistently slightly lower than younger subjects, suggesting greater indeterminacy. We predicted that these subjects would also be more likely to disprefer complex grammatical sentences than younger subjects, on the grounds of greater processing difficulty, but this was not confirmed. The test was not timed, so no data could be collected relating to reaction times which might have revealed some age-related variation in processing. Pye et al's (1992) GJ test, which was also not timed, produced a strong age effect only with much older individuals, in their 80s and 90s.

8.2.5 Comparisons with other Studies

The results of the present study can be compared with recent research in the same area. The studies which invite comparison are both those which compare adult and child or older and younger adult GJ test performance on wh-movement items and those comparing adult learners and native speakers. These studies mostly asked for nominal (ie absolute) judgements of acceptability or grammaticality; thus the findings express the closeness of subjects' approximation to linguists' judgements, not their preferences. In this study subjects using the magnitude estimation technique were asked to make judgements along three dimensions: grammaticality versus ungrammaticality; an ordering of types of (un)grammaticality; and ratios of ungrammaticality. Hence the present study is not always directly comparable with earlier work.

Our study clearly does not agree with Johnson and Newport's second study (1991), which involved Subjacency violations. Where they found a strong age of arrival effect amongst their Chinese subjects up to the age of 17 the present study does not. J&N (1991) found no evidence that adult arrivers judged Subjacency violations like native speakers, although they conclude that Subjacency "nevertheless survives in a weak and probabilistic form in adult learners" (237). Only the youngest child arrivers could match the performance of the native

speakers: a steady decline in accuracy (relative to native speakers) was charted from the age of 4 to 17. Moreover, J&N found that their adult Chinese subjects performed significantly above chance only on Relative Clause extractions, not on Noun Complement or Wh-island extractions. As has already been discussed, J&N's choice of aural presentation of test items depressed their test results (see Johnson 1992, Murphy 1997), and the proficiency level of their subjects has also been questioned (White and Genesee 1996). The present study used a written modality but also a slightly stricter selection criterion, and this may help to account for our different results. However, two aspects of their study are particularly relevant here. Their prediction that test performance would show not just an age effect but a critical period effect, is not borne out by their data: they found a native-speaker - non-native speaker divergence starting from the earliest age of arrival group (4-7) - although this only becomes significant with the second group (8-13). Test performance by comparison with native speakers worsened steadily over age of arrival up to and beyond adolescence. However, even the post-critical period arrivers show signs of "partial access" to the Subjacency constraint, and so they claim support for a long "sensitive period" rather than an abrupt critical period. However, although we did not attempt to compare pre-adolescent L2 knowledge at different ages, the present study found a fairly robust critical period effect for the parameterised value of Subjacency, and some signs of this effect also in the in the strength of preference figures for three of the Sentence Types.

The present findings also contrast with earlier studies which identified an unambiguous critical period for second language learning, notably Patkowski (1980).

White and Genesee (1996), who tested child and adult arrivers, found that, regardless of age of arrival, near native speakers of English performed like native speakers on their test. The main effect they identified was proficiency: there was some difference between the test performance of subjects independently assessed as lower in spoken English proficiency ("non-natives") and the near-natives, although less than in other studies. However, unlike the present study, the source of the difference was in the stronger violations: non-native subjects were less accurate in judging stronger Subjacency violations than weaker. This is surprising, not only in view of other studies (J&N 1991, Martohardojono 1992, Uziel 1993) but also because it strikingly conflicts with the theory: stronger Subjacency violations, especially for

speakers of L1s with wh-movement (most of their subjects were French native speakers) should be easier to reject even for speakers whose L2 is not near-native. Moreover, where our study found lower ratings for Relative Clause violations than for Noun Complement and Wh-island violations, for all subjects, suggesting that subjects were more confident about the unacceptability of the former items, we also found that adjunct extractions were rated high. White and Genesee found the greatest difference between near- and non-native speakers in this extraction domain too (F 6.71, $p < 0.01$). Adjunct extractions seemed to present the most problems for the less proficient (but still advanced) speakers. White and Genesee do not account for this result.

We hypothesized that, like Birdsong (1992), we would find not only child-adult differences in test results, but also differences between younger and older adult arrivers. Even on the strength of preference scores for Wh-islands, which were the only data to show clear age effects, an age effect *amongst* adult arrivers was not apparent. This finding conflicts with Birdsong (1992) who, in a rather different kind of study, found a clear age effect extending well into adulthood. In subsequent studies, Birdsong has also identified later adult arrivers as performing less accurately in GJ tests than earlier adult arrivers. Bearing in mind the reservations already expressed in regard to this study in Chapters 2 and 3, it can be noted that although Birdsong (1992) found a fairly strong overall age of arrival effect (r=.51, $p < 0.02$) and a stronger correlation for that-trace (ECP) violations (r=.60, $p < 0.005$), for "en"-Avant (a constraint on clitic movement, variously subsumed under Subjacency and the ECP - see Chapter 4) there was a negative correlation only with proficiency (r= -.67, $p < 0.001$), not with age of arrival[1].

8.2.5 Conclusion

Adult arrivers for the most part performed like native speakers and child arrivers. As has been said, it would only require subjects to transfer the L1 parameter setting of Subjacency to the L2 to achieve these results, as extractions out of Noun Complement, Adjunct and Relative clauses are barred in Italian as in English. Adults arrivers' performance on both types of parameterised Wh-island extractions suggests they are in fact using their L1 knowledge, as they found these sentences significantly more acceptable than did native speakers and child arrivers. Alternatively, are they rejecting both the L1 and the L2 value for

Subjacency? We did not investigate Italian speakers' intuitions about wh-movement in their L1 (nor have other studies), and this would be necessary to determine the acceptability in Italian, for Italian speakers, of the parameterised setting relative to invariant violations and to legitimate wh-movement. In the absence of such data, we may assume that our adult learners have adopted one of two strategies: if there is no difference in acceptability for Italian speakers between the finite parameterised setting and invariant legitimate wh-movement, then adult learners are treating English slightly differently and appear thus to have reset this parameter to an interlanguage value, neither Italian nor English. If finite parameterised sentences are of intermediate acceptability in Italian, then adult arrivers are simply transferring the Italian value to their L2. Rizzi (1982) identifies extractions from finite clauses as fully acceptable, but marks extractions from non-finite clauses "?" . Thus our subjects seem to have adopted the first strategy. In either case, the behaviour of the adult arrivers in this study was different from the child arrivers.

Do these findings support the claim that child L2 speakers have access to UG while adult access is inhibited to some extent? For most extraction types and across extraction domains, it is clear that L1 and L2 speakers of English have similar intuitions. What the strength of preference figures illustrate is that adult NNSs may be rather less confident than NSs about distinctions between controls and Subjacency violations. For the parameterised value of Subjacency alone there seems to be an age of arrival effect. We suggest that both child and adult learners have access to UG. Without UG, knowledge of wh-movement alone does not permit learners to recognise Subjacency and ECP violations. For the adult learners in our study this access seems inhibited by L1 knowledge. Hence, they have not been able to reset the Subjacency parameter to the English values, although they may have reset it to an intermediate value, hence still in obedience to UG.

J&N (1991) conclude that

> the older learner has ... a weakened or diminished set of universal constraints on human languages in general...The outcome of late second language learning, then, is an acquired language which, though still probabilistically similar to the target language, is imperfectly mastered and sometimes even violates universal constraints on human languages.(256)

The results of this, and many other studies, clearly contradict this conclusion. Our subjects showed no signs that their L2 grammar was only "probabilistically similar" to the target language. Nevertheless, the L2 grammar of the adult arrivers was not completely identical to the target grammar. More recent studies (Sorace 1993, Cook 1991) have suggested what an adult L2 grammar might look like. Sorace (1993) in particular has distinguished between incomplete and divergent representations of the L2 grammar, the former term referring to grammars that "fall short" of the target grammar, the latter to grammars "containing representations of L2 properties that are different from the native representations"(24). Without parallel data on our subjects' intuitions about the Italian setting of Subjacency, we cannot be sure, as stated above, whether the adult arrivers have transferred this setting to the L2, indicating an incomplete representation, or whether the degree to which they found these sentences acceptable in English indicates that they have developed a divergent (interlanguage) grammar, which nevertheless obeys universal constraints. One way in which such a grammar would diverge is that the adult learners might consider the "Italian values" optional in English, that is a possible variant of wh-movement constructions entailing no loss of grammaticality[2]. Hence the distinction most of them make between the grammatical control sentences and the "Italian" sentences.

8.3   An Acceptability Hierarchy for Wh-Movement Violations

8.3.1   Research Predictions

We predicted that native speakers and child arrivers would observe the grammaticality hierarchy specified by the "Barriers" account by according clearly differentiated scores to extraction types and extraction domains. What we found was that, across all groups, there were no significant differences between scores on ungrammatical extraction types, apart from the parameterised wh-island types (the "Italian values"), which have already been discussed. Strength of preference scores overall were lowest in the case of Adjunct Islands and highest for Relative Clause extractions, with Noun Complements and Wh-islands falling 2nd and 3rd highest. For extraction domains, where object extractions were expected to be judged more acceptable than subject extractions, the differences were not significant. This section will review possible explanations for the weakness, and the sometimes anomalous

character, of these results.

8.3.2 Comparisons with other Studies

Where recent studies have investigated L2 speakers' intuitions about the relative acceptability of different types of Subjacency violation, there have been contradictory results. Comparisons across studies are hampered by different choices of extraction types. On a pre-"Barriers" account, J&N (1991) predicted that their subjects would respond differently to parametrically varying violations than to invariant ones. However, they found that relative clause extractions were easier for subjects to reject than either CNP (invariant) or wh-island (parameterised) extractions. "Barriers" can account for this.

The "Barriers" account (described in Chapter 4) holds that grammaticality - hence acceptability - declines as the extracted element crosses more barriers. Therefore extractions from Wh-islands where one barrier is crossed in some languages but not in all (parameterised violations) are better than extractions from Noun Complements where one barrier is crossed in all (known) languages (invariant violation), while extractions from Relative Clauses and Adjunct Islands, where two invariant barriers are crossed, produce the worst effect. Martohardjono's results (1992) indeed confirm this. She investigated responses to wh-movement violations by subjects from one language with wh-movement (Italian) and two without (Chinese and Indonesian). She aimed to identify a pattern of relative acceptability by comparing accuracy rates on weak and strong types of Subjacency violation, and on object extractions versus subject extractions. All the language groups, and the native speaker controls, recorded a significantly different proportion of accurate responses for weak and for strong types. Indeed the Chinese and Indonesian groups, slightly less proficient, scored below chance level on weak types. As this study is directly comparable with the present experiment, it is worth inquiring into possible reasons for the very different results, especially as it was replicated, with similar results, by Uziel in 1993. In two respects Martohardjono (1992) and Uziel's (1993) experimental design differed from the present study: the test was administered aurally, and the test items were constructed in such a way as to minimise lexical differences between violation types, and between grammatical and ungrammatical sentences. In addition, all moved elements were objects or subjects - there were no sentences with moved adjuncts. These two features of the test materials may have

served to make syntactic differences more salient. All moved wh-elements were of the "which X" variety - no other question words were used. Kluender (1992) argues that the more referentially specific a moved NP, the more interpretable and the more acceptable it becomes; a moved object NP taking the "which X" form may therefore be more acceptable than "what" in an ungrammatical sentence:

(1) *Which soup did the man leave the table after the waiter spilled?
(2) *What did the man leave the table after the waiter spilled?

The native speakers of English in Martohardjono's and Uziel's studies are notably less accurate on object extractions from Wh-islands and Noun Complements than in other studies, scoring around 75% compared to around 90% (see J&N 1991, Schachter 1990, White 1988, White and Genesee 1996). It is possible that this wide gap between one and two barrier violations is at least partially an artefact of their test design.

As already noted, White and Genesee (1996) also measure accuracy rates of different types and strengths of ungrammatical wh-movement. In contrast to Martohardjono (1992) and Uziel (1993), they found that proficiency affected accuracy rates on stronger violations (extractions from Relative Clauses and Adjunct Islands)[2]. Their non-native subjects were more likely to judge as acceptable sentences like (3), where an object is extracted from an adjunct clause than sentences like (4), where an object is also moved but where the extraction site is located in a noun complement:

(3)* Who did you meet Tom after you saw t?
(4)* What did you hear the announcement that Ann had received t?

8.3.3 Alternative Explanations

The advantage of magnitude estimation over dichotomous scoring should be that it allows subjects to make fine distinctions between levels of grammaticality. Results in the present study show that subjects did distinguish between violation types, and between object and subject extractions - that is, between Subjacency and ECP violations - but differences in

ratings failed to reach significance. In particular the anomalous ratings for Adjunct Islands and for extractions of adjuncts from other Sentence Types require an explanation. As the theory clearly predicts a hierarchy which is not evident in the present data, it is necessary to look elsewhere for factors that might account for these data. In the first place, we will look for weaknesses in the research design. It may be that subjects were not judging test sentences on syntactic grounds alone, and that the very weak differences between their judgements of different violation types reflect the interference of other factors. We examine here three possible accounts of subjects' behaviour in the experiment.

A. Correctability

The absence of significant differences among extraction types within Sentence Types, and among Sentence Types themselves strongly suggests that subjects have not judged the test items only - or primarily - on syntactic grounds. In spite of the instructions, which made it clear that interpretability was not an issue, subjects seemed to have - at least in some cases - judged sentences for coherence. For example, among the Subjacency violations for Adjunct Islands native speakers judged (5) to be twice as good as (6):

(5) *Which pub did George give his lecture after he left yesterday?
(6)* What do you wonder where Fred will want to move to after selling?

Although the result is slightly implausible, (5) can be easily improved by inserting "in" before the moved element (giving "In which pub..."), while correction is much more difficult in (6). Crain and Fodor (1987) have shown that correctable constraint violations (in their case hinging on a single word, as in (5)) are judged quite differently from uncorrectable violations, although they may be equally ungrammatical. One reason at least for the discrepant figures for Noun Complement Subjacency and ECP violations might be the greater interpretability and correctability of (7) - an ECP violation - compared to (8) (Subjacency):

(7)* Who have you heard the news that has got the job?
(8)* Who did they spread the rumour that the headteacher had punished?

B. Parsing Strategies

The unexpected results for adjunct extractions, whereby ungrammatically extracted adjuncts, in all Sentence Types, were generally rated higher than object or subject extractions, against theoretical predictions, and unlike empirical research findings (Martohardjono 1992 and Uziel 1993) requires explanation. A factor that may be relevant here is the way subjects parsed sentences. The scores for sentences like (9) are higher than for sentences like (10), suggesting that subjects may have parsed them differently:

(9) * Where did Ann want to go home after drinking a bottle of vodka with her friends?

(10) * How long does Dr Smith believe which problem his students can solve after working on it?

The Minimal or Local Attachment Principle (Frazier and Fodor 1980) says that "of two possible ways of attaching a node to a partial phrase marker, the simpler will be preferred. The simpler attachment is the one with fewer nodes intervening between the new node and the existing phrase marker" (quoted in Garnham 1984). The moved wh-element in (9) could thus be attached to the higher VP, resulting in a sentence that was acceptable, ie plausible and interpretable, if slightly lacking in coherence:

(11a) Where did Ann want to go home t2 after drinking a bottle of vodka with her friends t1?

where t1 represents the intended extraction site, and t2 a semantically plausible site. Compare the shorter:

(11b) Where did Ann seem happy after going t?

In (11b) "seem happy" in place of "go home" discourages attachment of the moved adjunct to the matrix verb. The gap left by the moved adjunct is also more salient, perhaps because of the shorter distance between filler and gap.[4]

The gap left by a moved subject is more salient than the gap left by other moved elements. Juffs and Harrington's (1995) study (discussed in Chapter 4) showed that the asymmetry

between subject and object wh-extraction can be explained by Pritchett's Generalized Theta Attachment: every principle of the Syntax attempts to be maximally satisfied at every point during processing (Pritchett 1992:138).

Parsing subject gaps requires more reanalyses than parsing object gaps, and the gap left by a moved adjunct does not require reanalysis either, as in (12):

(12 Where did Jane believe her friend went _ ?

Juffs and Harrington (1995) make use of Generalized Theta Attachment to explain the poorer performance of Chinese subjects on a Subjacency test compared to native speakers, and show that in a timed test the former took longer to read subject gaps than object gaps in both grammatical and ungrammatical sentences. Although the absence of wh-movement in Chinese partly accounts for the greater difficulty in parsing found by these subjects, native speakers too took longer to parse subject gaps - ie they also needed to reanalyse a subject gap but not an object gap. Adjunct gaps do not require reanalysis while the sentence is parsed - they can be treated like object gaps. It may be that certain ungrammatical adjunct extractions, where the trace is at the end of the sentence, appear, like object extractions, less deformed than subject extractions, especially in the case of longer sentences.

If this was the case here, the claim must not be pushed too vigorously: Hawkins (1994) argues persuasively for a processing account of acceptability hierarchies precisely because processing accounts (his Early Immediate Constituents Principle) can explain the same behavioural patterns as can syntactic principles - not contrary ones. It seems most likely that if processing or parsing is responsible for at least some of the adjunct results, it is due to the test sentences themselves, not to their violation type. In addition, many of the adjunct extractions were rated even higher than object extractions, which renders this argument less convincing.

C. A Syntactic Explanation

A third possibility derives from asymmetries between subject extraction and adjunct extraction. In the following pair of sentences, (13a) was rated as significantly worse than

(13b) by all subjects.

(13a)* Who do you know whether they said, in the hotel lounge, moved to Ireland?

(13b)* At what age do you know whether John said he moved to Ireland?

Both sentences represent ECP violations. In (13a), an argument is extracted from subject position of an embedded question, while in (13b) an adjunct is moved from the same kind of clause (it could also be read as moved from the higher IP "said", but this should not affect the outcome). As several writers have shown, although both adjunct and subject movement are subject to the ECP (Huang 1982) there are asymmetries (Rizzi 1994). For example, adjuncts can be moved over overt complementizers, but subjects cannot[5]. Also, as Rizzi (1994) has shown, different kinds of adjuncts adjoin to different nodes in the structure, and this affects their movement possibilities. Rizzi argues that sentence adverbials like "why" can be directly base-generated in Comp, so do not move, while manner adverbials (and probably time and place adverbials) are adjoined to VP, are properly head-governed by IP (or TP), and so are moveable. Extraction of adjuncts is thus partly subject to different constraints to that of subjects, and may, together with lexical considerations result in varying degrees of acceptability. Nevertheless, both Rizzi (1994) and Culicover (1996) agree that on the whole adjunct extractions result in worse rather than better sentences, compared to subject extractions.

In the adjunct extraction items in the present study, the extracted adjuncts were mainly time and place adverbials, as "why" can too readily be parsed as attached to a matrix verb of cognition and "how" is in free distribution with "why" in many Scottish dialects. Thus it is not possible to make a post-hoc comparison of different types of adverbials in adjunct extractions.

To conclude, it seems most likely that some of the test items allowed subjects to correct them by making small changes and thus render them more acceptable than was intended. Others may have been in some way "parsed into acceptability" by moving the extraction site from the lower to the higher clause. The length of the items, built in to the design of the test items in the hope of testing subjects more stringently, may have combined with some lexical

choices to encourage this. Both correctability and the Minimal Attachment Principle seem to have adulterated the reading of adjunct extractions more than others, although some subject extractions were also rated higher than expected, perhaps for similar reasons.

To try to discover how subjects were reading the long adjunct extractions, and to what extent sentence length, correctability and parsing strategies were affecting judgements, a small-scale comprehension test was devised. It was hypothesized firstly that adjunct extractions would be more sensitive than object extractions to the Minimal Attachment Principle, and that they would therefore tend to be parsed as though the extraction site were located in the upper, matrix clause. Secondly it was predicted that where minor lexical additions and deletions would appear to improve the grammaticality of items, subjects would be more willing to do this in the case of adjunct extractions than where arguments were extracted. Seven of the ungrammatical adjunct extraction sentences, with shorter versions and grammatical adjunct extraction items were presented to 3 English speakers and 3 highly proficient Italian speakers, none of whom had taken part in the original study. Subjects were asked to rewrite or correct 16 grammatical and ungrammatical sentences and to grade each corrected sentence for ease of correction on a scale of 1 (=very easy to correct) to 5 (= very difficult to correct). The procedure was adapted from Crain and Fodor (1987).

There were no differences between the performances of native and non-native subjects. The corrections were scored for grammaticality, interpretation and identification of extraction site(s). Total scores for all subjects ranged between 27 and 32 out of a possible 41, apart from one native speaker who scored 23.5.

Results of this experiment revealed that object extractions proved interpretable: that is, subjects correctly identified the extraction site in sentences like "who do the doctors know whether the medicine cured?" and "What did my daughter phone the police on the day when someone stole?", although they reported that the latter sentence was considerably more difficult to process (mean difficulty ratings 2.2 and 3.4). However, in the case of adjunct extractions, subjects were much more likely to make changes to the test items by adding or deleting function words, in order to force extraction from the matrix clause in preference to long-distance movement. Thus the two items which exemplified adjunct extraction from

adjunct islands "Where did Ann want to go home after drinking a bottle of vodka with her friends?" and "For how much do you wonder which part of the city Fred will choose to move to after selling his flat?" were found relatively easy to correct (mean difficulty ratings were 2.8 and 2.6); however 5 subjects identified the wrong site in the former and 4 in the latter. Subjects added "to" after "home" in the former sentence, thus locating the trace in the matrix clause; in the latter sentence, subjects deleted "for how much", producing sentences like "Which part of the city will Fred choose to move to after selling his flat?"; although 2 subjects did correctly interpret the sentence.

Similarly, in the case of adjunct extractions from an NP complement and a Wh-island, subjects added or deleted words in order to parse the wh element with the main clause: "Where did they wonder whether their uncle had sent the parcel?" was corrected to "Where did they wonder had their uncle sent the parcel?" and " Did they wonder whether their uncle had sent the parcel? " Slightly shorter items instantiating ungrammatical adjunct extraction and where the choice of lexis discouraged incorrect parsing still posed some problems: for "Where was Fred worried about the house he bought", 2 out of 6 changed "where" to "why", although the others interpreted the sentence correctly (eg "Was Fred worried about where the house he bought was?").

J&N (1991) also included a Subjacency comprehension test, which they administered to a subset of their adult Chinese learners of English and native speakers. Their purpose was to discover whether or not the former were parsing the Subjacency violations correctly (as illegitimate extractions from embedded clauses rather than legitimate extractions from matrix clauses). The results of this small test showed that the Chinese adults were indeed interpreting the violations correctly - and in fact were able to interpret them more successfully than the native speaker controls.

It would appear that ungrammatical adjunct extractions present greater difficulty for interpretation than object extractions, but also greater opportunities for *re*interpretation, for both native and non-native speakers. The hypotheses given above therefore seem to be confirmed by this small and informal experiment. Adjunct extractions have rarely been included in Subjacency and ECP studies of second language speakers, and we can perhaps

see why.

8.3.4 The Use of Magnitude Estimation

It has already been remarked that the greater freedom in the making of grammaticality judgements allowed by magnitude estimation seems to lead to naive informants giving rather different rankings of violations from linguists. Less than "perfect" accuracy from native speakers in GJ tests is of course to be expected, but the magnitude estimation technique seems to highlight such divergence. Bard et al (1996), in their small-scale exercise undertaken in order to make such a comparison found that although all their subjects tended to agree on the top and bottom levels in the hierarchy, there was considerable disagreement on intermediate levels. Thus, their 4 linguistically untutored anatomy undergraduates ranked an ECP violation (an adjunct extraction marked "* *" in Haegeman 1994) much higher than a Subjacency violation (an object extraction marked only "?" in Haegeman 1994). The experienced linguists also found a weak Subjacency violation less acceptable than an ECP violation (also an adjunct extraction), but better than an alternative lexicalisation (which discouraged a higher attachment of the moved adjunct). They make the point that lexicalisations can influence the judgements of even experienced linguists, so "averaging over different lexicalisations may prove necessary" (48) - as was done in the present study. Bard et al (1996) maintain that the advantage of magnitude estimation lies in the greater delicacy and robustness with which it distinguishes among linguistic categories. This entails a concomitant risk perhaps, in that problematic lexicalisations or semantic/pragmatic features may attract attention in a way that nominal judgements avoid. It may also be the case that ME is less suited to tests involving Subjacency and ECP violations than to studies where - as in Sorace (1993) - semantic and syntactic factors are intertwined. That is, conceptualizing "preference" in relation to levels of grammaticality alone, where interpretation can be equally obscure at all levels, may be too difficult for naive subjects. If more training is required, this begs the question as to the value of using non-linguists as informants.

8.4 Conclusion

Problems with the research design have made interpretation of the results of the experiment

more difficult than it should have been. The design of some of the test sentences allowed subjects to - by one means or another - correct or parse them so as to be less unacceptable than the experimenter intended. In addition, finding subjects at near-native levels of proficiency for all age of arrival groups or subjects at two levels of proficiency for all groups limited the scope of the study.

Nevertheless, we have found fairly convincing evidence that advanced L2 speakers from a wh-movement language like Italian, long-term residents in the L2 community, recognise invariant Subjacency violations and ECP violations, regardless of age of arrival and regardless of proficiency. We conclude that the data strongly suggest that adult learners are able to access UG when acquiring a second language. The fact that in the present case the L1 and the L2 are configured similarly for Subjacency, for the most part, does not mean that learners can recognise Subjacency violations in English simply by referring to their knowledge of Italian: the level of their proficiency in English is crucial in that without a near-native knowledge of English wh-movement, especially of long movement, learners would be, like the Spanish learners in Johnson (1988), unable to display a near-native knowledge of invariant values of Subjacency.

For the finite parameterised value of the Subjacency Condition, there is an age of arrival effect. Adult learners seem unable to reset the parameter. We also found that, where strength of preference figures show variation amongst subjects this allows us to discriminate between child arrivers and adult arrivers, but not between adults with different ages of arrival. We conclude that adult arrivers are more likely to transfer the L1 setting than child arrivers, which in turn confirms our claim that UG is indeed accessed by adult learners but that this access is mediated in some way by prior L1 knowledge.

This study does lend some support to the critical period hypothesis. However, the results presented here do not support the findings of Birdsong (1992) and Birdsong and Molis (1997) on a *continued* age of arrival effect stretching into adulthood. Although it less likely that older adults will achieve the same proficiency as younger adults or children, it is not impossible, and, as mentioned above, lower oral proficiency may not necessarily translate into less native-like judgements of ungrammatical sentences.

[1] White and Juffs (1998) speculate that the slight advantage displayed by their "China group" over the "Canada group" in their experiment is attributable to the fact that the former were younger at their first communicative contact with English (average age for China group:22.8; average age for Canada group: 32.25). But Birdsong's effect, which they claim to find support for, was seen only in individuals with an age of arrival after their mid 30s.

[2] Optional constructions, such as extraposition and heavy NP shift in English or auxiliary change with some modals in Italian are constructions where either of two alternatives is equally acceptable. Thus, in English: "I read a review of John's book last week" and "I read a review last week of John's book" are both acceptable by native speakers (Fukui (1993)). In Sorace (1993) Italian respondents rated "Maria non ha potuto venire alla mia festa" (Maria couldn't come to my party) and "Mia figlia non e potuta venire a scuola " (my daughter couldn't come to school) as equally acceptable.

[3] Following Cinque (1990) White and Genesee (1996) take Noun Complements to be strong islands for Subjacency.

[4] In "Ann went to London where she saw an exhibition with John", the interpretation "saw an exhibition with John" is preferred to "went to London with John" by the Local Attachment Principle.

[5] Compare (1) Who did you say *that/θ left yesterday, and (2) When did you say that/θ John left?

CHAPTER 9

CONCLUSION

9.1   What is the Age Effect for Adult Learners?

9.2   UG: Adult Access or Not?

9.1   What is the Age Effect for Adult Learners?

We began this study by asking whether the age at which an adult is first exposed to a second language has a bearing on their final mental representation of the second language grammar. Although this question has been thoroughly investigated in relation to child-adult differences, the issue of an age factor in adulthood has been somewhat neglected in SLA research. The age factor can "kick in" at many different life stages, not just in the early years, and it seems surprising that age of arrival in adulthood has attracted so little attention. The many studies of the linguistic abilities of older adults in relation to their first language suggest, although not unanimously, that the decline of short-term memory is the main cause of changes in language use in later life. Processing and comprehension - ie "performance" - change quite significantly over the lifespan, but language knowledge or "competence" is maintained virtually unchanged. Even stroke or other neural insult which may cause partial or temporary loss of language, cannot - it seems - destroy the language instinct. On the other hand, the handful of cases of late acquisition of a first language, whether verbal or signed, strongly suggest that this is bound to be incomplete if begun after puberty and the offset of the critical period, even though the neurological bases are still largely unknown. Nevertheless, once acquired, it seems that language is with us for life.

There is strong evidence that the learning of a second language is constrained by the same maturational processes, although to a lesser extent, as first language acquisition. Thus, pre-pubescent exposure to L2 mostly results in near-native attainment. However, post-critical period L2 acquisition is not only possible but leads *in most cases of extensive exposure* to an ability to communicate satisfactorily in "everyday" domains. Thus there are few adults who have lived in an L2 environment for more than 2 or 3 years who cannot "get by" in such domains. Schmidt's study (1981) of the fossilized English of his Japanese subject, Wes, gives striking testimony of the human ability to communicate in the face of severely limited language learning skills.

A very small proportion of adult acquirers can attain the proficiency levels quite normally achieved by early learners, and it has been suggested (Schneiderman and Desmarais 1988, Novoa, Fein and Obler 1988, Smith and Tsimpli 1991, Ioup et al 1994) that unusual brain development, neurocognitive flexibility or "abnormal" distribution of functions between the

hemispheres can account for this - ie not an exceptionally assimilative psychological profile nor greater cognitive abilities. However, the adult learners we have been concerned with in this study are not exceptional in this sense, although they are a minority of all adult L2 speakers. The adult subjects selected for the experiment all retained Italian accents and most betrayed occasional signs of fossilization in their English production. They had all been living in the L2 environment for many years, and most had used English in professional and working situations. They can thus be more accurately described as "expert non-native speakers".

Studies discussed in previous chapters suggest that there are many near-natives and expert non-natives who have intuitions about certain features of the target grammar that are identical or almost identical to those of native speakers. The conclusion to be drawn from these studies is that a native-like representation of the L2 grammar *can* be achieved by adult learners, and not only by the most gifted. However, their results and also the results of the present experiment are also consonant with the argument that while certain *core* properties of the L2 grammar can be completely acquired by such adults, peripheral and language particular features may be beyond their grasp. Notwithstanding the constructions that may be placed on such results, they offer powerful evidence against a critical period for second language learning. If the critical period effect can be overridden by a substantial minority of adult learners, it clearly lacks the power it exerts in first language learning. Something like the "exercise hypothesis" may fit the evidence better. This predicts

> differential success for child and adult FL learners, but equal success for child and adult SL learners. FLA by adult starters, eg the learning of American Sign Language (ASL) by congenitally deaf adults, will either by inaccessible (strong version) or less successful than child FLA (weak version), because it will be irregular and/or incomplete. SLA, on the other hand, will be unaffected by the age of the learner, according to this view, since the language learning capacity remains intact once activated. Long (1990: 254-55)

Recall that Mayberry and Eichen (1991) found their adolescent second language signers more proficient than late first language signers. Some of the most recent neurological research on the bilingual brain (Kim et al 1997) tells us that for later bilinguals - older children in this case - the area of the brain that processes heard utterances, Wernicke's area, is not localised for the two languages in the way that Broca's area for speech production is. A single processing function may serve both languages, while later acquisition of a second

language may require the activation of a separate function for speech production. This finding, especially if confirmed for adult bilinguals, may be related to the phenomenon discussed above, and in other studies, in which adult learners with less than native-like production are nonetheless able to make native-like metalinguistic judgements.

In assuming that L2 learners do reach asymptote after a certain period of exposure such as 3 or 5 years in the L2 community, we enable comparisons to be made between them and native speakers that may not be primarily based on measured proficiency and may not be empirically verifiable. We are not thereby characterising either the proficiency or the underlying competence of the non-natives. As was discussed in Chapter 3, making meaningful, quantifiable comparisons between the proficiency of such learners and native speakers is difficult and risky. However, it can be assumed that both groups have steady-state grammars of the target language. Intra-subject differences should therefore be constant, and should say something about *possible* convergences or divergences between native and non-native grammars. What was found was that both adult and child arrivers with long exposure and who can be termed expert speakers converged with native speakers on core aspects of Subjacency and the ECP, but diverged on peripheral features. Thus, for the most part, the steady state of both groups as regards knowledge of wh-movement was the same, and even the less proficient non-natives within this sample did not diverge significantly. However, on the finite parameterised value of Subjacency, adult arrivers had clearly not restructured their knowledge in accordance with the L2, while child arrivers had.

The comparison of adult learners with different ages of arrival failed to reveal that age of arrival was a factor in metalinguistic behaviour after adolescence. Had a significant age-related decline in the ability to match NS judgements on the part of the adult learners appeared, we would have been able to confirm Birdsong's (1992) findings. This was not the case: our findings tend rather to support those of White and Genesee (1996).

However, had our results been different, it might have been a puzzle to explain them. Birdsong (1992) attributes the age-related decline shown in his data to "maturational effects" (736), without further explanation. In fact, what his data seem to point to is a *much later offset time* for the critical period, rather than a continuing decline through adulthood. "If the performance of exceptional learners is to be accounted for in biological terms, then the hypothesized end of the critical period must be pushed well past puberty, or the 'window of

opportunity' for language learning must be extended and made flexible"(742). In his replications both of Coppieters (1987) and of Johnson and Newport (1989) (Birdsong 1989, Birdsong and Molis 1997), it is clear that the early thirties marks a boundary for proficiency, although numbers of subjects who arrived after their mid-thirties are very small (n=5, n=4 respectively). However, there seems to be no *maturational* reason for fixing the end of the critical period at 34, and none is suggested. Perhaps what Birdsong should be arguing for is a type of skilled or expert learner for whom the critical period is irrelevant but in whom ageing, not maturational, effects, may appear towards middle age. Hence this learner never loses the brain function that allows complete language learning (in Hurford's metaphor, the light is kept on), but is unable to resist the effects of short-term memory decline. While we were unable to find such late arrivers for the main study, in the pilot study the oldest arriver (aged 40) did show the greatest amount of divergence from the NS norm, while being rated relatively highly on proficiency.

"Maturational effects" amongst 30 year-olds are hard to imagine. Therefore, if there are late age effects amongst near-native speakers, it seems much more likely that psychological or memory factors are at work: short-term memory tends to be reduced at least during the learning of a second language (Brown and Hulme 1992, Harrington 1992) and the kind of decrements noted from middle-age on in studies of ageing may combine with this to impact more on later L2 learners. Hurford (1996) proposes that the particular neurological function that enables a first language to be acquired during the critical period (but not later) is the "phonological loop" subcomponent of working memory. It appears that this subcomponent is much less significant in parsing than in acquisition: patients with short-term memory deficit are only impaired on processing relatively complex sentences, but there is evidence that such patients (both children and adults) find considerable difficulty in learning novel vocabulary. We suggest that the kind of short-term memory deficits found in normally ageing populations may be precisely the factor that inhibits second language acquisition, even for apparently above-average learners, from middle age on.

One direction future research in this field might take, therefore, is to investigate longitudinally the role of memory in late L2 acquisition, focusing on the process older learners go through compared to younger adult learners. Selecting learners with a range of aptitudes and rates of learning could also shed light on the role of individual differences in working memory. Studies such as Ioup et al (1994) are inevitably historical, in that their

language learners *turned out to be* talented. It should not be impossible to identify talented language learners early on and to compare memory capacity with both performance and competence in the L2

The findings on L2 proficiency in the study are problematic. Even Birdsong (1992) found some effect of proficiency for some results, although most of the variation in his results is explained by age of arrival. We have pointed to the difficulty of assessing proficiency objectively at this level. It has also been argued that long exposure to the L2, although it may not result in native-like performance, may nevertheless develop native-like intuitions about the L2 grammar. Clearly, more research into the cluster of factors that produce this effect is called for.

Difficulties in finding enough late adult acquirers at native-like proficiency levels seem likely to beset this kind of study forever; quite simply, people who are interested in and capable of achieving such high levels of language performance settle in the L2 community in their 20s, not later.

9.2   UG: Adult Access or Not?

The results of the present study strongly suggest that adult learners have access to Universal Grammar in acquiring a second language, in line with many other studies. Therefore adults are not required to apply general cognitive processes to this task. However, the weaker achievement of adults strongly suggest that they do not start with the principles and parameters of UG set at default values, as children probably do in both first and second language acquisition. The hypothesis that most closely fits the data presented here is that which proposes that adult access to UG is mediated by L1 knowledge. Hence, adults start by assuming L1 settings or values, and reset parameters in accordance with the L2 input, *as far as they can*. We suggest that our data support the Full Transfer/Full Access model of Schwartz and Sprouse (1996) who argue that "all the properties of the L1 computational system transfer as a block" (66-67). If only positive evidence is available for restructuring, aspects of the target grammar that are impervious to positive data will fossilize. If Full

Transfer/Full Access also applies to child SLA, which is not ruled out, then an explanation for the more efficient operation of positive evidence in the child must be sought. As we have shown in Chapter 4, English is more restrictive in regard to Subjacency than Italian, and is therefore in a subset relation to it. According to Subset theory, the subset grammar is the unmarked option, and positive input can only force restructuring from an unmarked to a marked grammar, not the other way round. Hence English learners of Italian should accept (1) (from Rizzi 1982), but Italian learners of English do not, as we have seen, reject (2).

(1) La nuova idea di Giorgio di cui immagino che cosa pensi diverrà presto di pubblico dominio.

(2) The students dislike the course which I can't understand why the college organised.

On the other hand, the FT/FA hypothesis proposes that when knowledge of UG interacts with the L2 input, some parameters cannot be reset to the L2 value, but equally do not remain at the L1 setting: input thus forces an interlanguage setting which is still constrained by UG. In the present study, we were unable to present the evidence that would show that the non-native subjects had reset the Subjacency parameter in this way: only an investigation of the status of the finite "Italian value" in native grammars would reveal whether it is completely acceptable or only marginal. Given that the adult learners accorded this value marginal acceptability in English we have suggested that our adult subjects show evidence of an incomplete, rather than a divergent representation of English. Recall that the French speakers of Italian in Sorace (1993) changed an optional choice of auxiliary in restructuring constructions into a categorical rule, thus overgeneralizing, while the English speakers appeared to ignore the evidence for auxiliary change in these cases. Sorace (1993) characterises the French speakers as having divergent and the English speakers incomplete grammars of Italian, and suggests that selective use of the available input is to blame. It can only be the L1 grammar that induces such selective attention: for unaccusative verbs, Italian is the subset language and rules for auxiliary change for these verbs are more restrictive for Italian than for French. In the present study, the Italian speakers of English appear not to have noticed evidence in the input banning the Italian setting of the Subjacency Condition in English , ie the absence of sentences instantiating the broader Italian value. According to the Subset Principle, they *could* not notice this, as Italian is in a superset relation to English with respect to this feature. Absence of positive evidence that English is more restrictive than Italian as regards the Subjacency Condition can account for divergent rather than incomplete

L2 grammars. Recall however that Cook (1991, 1992, 1995) hypothesized a bilingual competence, or multi-competence, wherein both languages may be represented differently from the respective monolingual competences: "the L1 and L2 systems are symbiotic within the same mind, as is clearly evident, for example in the complex phenomenon of code-switching" (Cook 1995: 57). It is not ruled out by our data that the Italian speakers of English have divergent representations of both Italian and English, when compared to the grammars of monolingual speakers.

The findings presented in this study also appear to support a weak version of the "interference hypothesis" proposed by Hurford (1991): the learning of a second language is inhibited to some extent by prior learning of a first. But if this hypothesis also predicts that the older the learner, the lower the level of ultimate attainment - as Hurford seems to argue - then our findings clearly fail to confirm it.

The question of fossilization at high levels of proficiency remains. Whether it is the effect knowledge of the L1 has on the accessibility of some properties of the L2, as proposed by the Full Transfer/Full Access hypothesis, or the subset-superset relation of the two languages, it nevertheless appears that some - a very small number - of adult learners overcome these barriers, while most cannot. Studies of talented learners have indicated that unusual brain organisation is responsible; however, it also seems clear that learning strategies can make a difference. Julie in the study by Ioup et al (1994) *noticed* aspects of Egyptian Arabic which most naturalistic learners would miss. Studies of learning styles, such as Leow (1997) and Robinson (1995) have claimed to support Schmidt's "noticing" hypothesis, (Schmidt 1990,1993, quoted in Leow 1997) which advantages adult learners who are able to pay attention to form. The problem is how to operationalise noticing or "awareness". Leow (1997) defines awareness or noticing as a behavioural or cognitive change in response to a subjective experience or external stimulus. Using think-aloud protocols, recorded as subjects completed a crossword puzzle, in pairs, he demonstrates that subjects (low level English learners of Spanish) who register awareness of a rule – in this case, regular tense forms - are significantly more likely to produce accurate responses. However, the direction of causality is unclear, and whether "meta-awareness" is a strategy that can be taught, or a feature of skilled language learners, remains in doubt. If the latter, the argument becomes circular – good learners are good learners. Think-aloud protocols, it will be recalled, were employed by Coppieters (1987), and criticised subsequently for distorting metalinguistic judgements.

However it could perhaps be utilized to investigate whether noticing or metalinguistic awareness is indeed what distinguishes fossilised from non-fossilised learners.

Johnson et al (1996) find "at the center of most discussions of child-adult differences in language acquisition" (336) the question: "is [..] the older learner's grammar... *qualitatively* different from that of the young learner?" They argue that the adult learner's grammar is made up of elements of the L2 grammar, elements that belong to a non-target interlanguage grammar, but also "areas outside of the adult learner's grammar in which s/he lacks determinate knowledge and which are therefore subject to a variety of probabilistic, nondeterministic response factors, for example response biases, practice effects and guessing." (347). In the present study however we found no more evidence of guessing on the part of the non-native speakers than for the native speakers (although Johnson et al's research is calculated specifically to search out guessing, unlike most ultimate attainment studies). The subjects selected for our study had no difficulty in distinguishing violations from grammatical wh-movement sentences, and gave *determinate* responses in the single area where they differed from native speakers. It can be argued that as with previous research using Johnson and Newport's (1989) testing instrument, modality, and the absence of proficiency screening can account for at least some of the results in Johnson et al (1996).

Future research should investigate the status of constructions of the type of (1) and (2) not only in the L2 grammars of Italian speakers of English, as here, but also in the L2 grammars of English speakers of Italian, and in the L1 grammars of Italians. Such research would reveal more about the nature of the L2 competence in near-natives who started to learn their second language as adults.

*

REFERENCES

Asher, J and Garcia, R (1969) The optimal age to learn a foreign language. Modern Language Journal 53: 334-341

Baddeley, A. Papagno, C. and Vallar, G (1988) When longterm learning depends on short-term storage. Journal of Memory and Language 27: 586-595

- and Hitch, G. (1974) Working memory. In Bower, G (ed) The Psychology of Learning and Motivation volume 1. New York: Academic Press

- (1998) Recent developments in working memory. Current Opinion in Neurobiology 8:234-238

Bard, E, Robertson, D. and Sorace, A. (1996) Magnitude estimation of linguistic acceptability Language 72,1:32-68

Berquist, B. (1997) Individual differences in working memory span and L2 proficiency: capacity or processing efficiency? Paper presented at GALA 1997

Berwick, R.C. (1985) The Acquisition of Syntactic Knowledge. Cambridge, Mass: The MIT Press

Bialystok, E (1997) The structure of age: in search of barriers to second language acquisition. Second Language Research 13,2: 116-137

- and Sharwood-Smith, M. (1986) Interlanguage is not a state of mind: an evaluation of the construct for second language acquisition. Applied Linguistics 6:101-117

Birdsong, D. (1989) Metalinguistic Performance and Interlinguistic Competence. Berlin: Springer-Verlag

- (1992) Ultimate attainment in second language acquisition. Language 68:706-755

- and Molis, M (1997) Dissociating age and maturation in L2A: a replication of Johnson and Newport (1989). BUCLD 1997

Blanche, P. and Merino, B. (1989) Self-assessment of foreign language skills: implications for teachers and researchers. Language Learning 39: 313-340

Bley-Vroman R W, Felix, S.U. and Ioup, G. (1988) The accessibility of Universal Grammar in adult language learning Second Language Research 4:1-32

Bloom, R.L., Mullins, J and Paternostro, P (1996) Changes in processing adverbial conjuncts throughout adulthood. Applied Psycholinguistics 17: 105-116

Bloomfield, L. (1933) Language. New York: Holt, Rinehart and Winston
Borer, H and K Wexler (1987) The maturation of syntax. In Roeper, T and

Williams,E. (eds) Parameter Setting, Dordrecht: Reidel

Bornstein, M.H. (1989) Sensitive periods in development: structural characteristics and causal interpretations. Psychological Bulletin 105:179-197

Bourdages, J. (1992) Parsing complex NPs in French. In Goodluck, H and Rochemont, M. (eds) Island Constraints: Theory, Acquisition and Processing. Dordrecht:Kluwer

Brebion, G., Ehrlich, M.F. and Tardieu, H. (1995)Working memory in older subjects - dealing with ongoing and stored information in language comprehension. Psychological Research - Psychologische Forschung. 58,3:225-232

Brown, G.D.A. and Hulme, C (1992) Cognitive psychology and second language processing: the role of short-term memory. In R.J.Harris (ed) Cognitive Processing in Bilinguals Amsterdam: Elsevier Science Publishers

Carroll, J M, Bever, T.G. and Pollack, C.R.(1981) The non-uniqueness of linguistic intuitions Language 57: 368

Case, R., Kurland, D.M. and Goldberg, J (1982) Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology 33:386-404

Cazden, C.B., Cancino, H., Rosansky, E.J. and Schumann, J.H. (1975) Second language

acquisition sequences in children, adolescents and adults. Final Report, National
    Institute of Education.

Chaudron, C (1983) Research on metalinguistic judgments: a review of theory, methods,
    and results. Language Learning 33:343-377

Cheung, H. (1996) Non-word span as a unique predictor of second-language vocabulary
    Learning. Developmental Psychology 32,5: 867-873

Chomsky, N (1965) Aspects of the Theory of Syntax Cambridge, Mass: MIT Press

-    (1981) Lectures on Government and Binding. Dordrecht: Foris

-    (1982) Rules and Representations. New York: Columbia University Press

-    (1986a) Barriers. Cambridge, Mass: The MIT Press

-    (1986b) Knowledge of Language. New York:Praeger

-    (1989) Some notes on economy of derivation and representation. MIT Working Papers
        in Linguistics 10:43-74

-    (1995) The Minimalist Program. Cambridge, Mass: The MIT Press

Cinque, G (1990) Types of A-Dependencies. Cambridge, Mass: MIT Press

Clahsen, H (1988) Critical phases of grammar development: a study of the acquisition of
    negation in children and adults. In Jorden, P. and Lalleman, J. (eds) Language
    Development. Dordrecht: Foris

Clahsen, H and Muysken, P. (1986) The availability of Universal Grammar to adult and child
    Learners - a study of the acquisition of German word order. Second Language
    Research 2: 93-119

Colombo, J (1982) The critical period concept: research, methodology and theoretical issues.
    Psychological Bulletin 91:260-275

Cook, V (1989)Word order and timed grammaticality judgements. Paper read at LARS
    Conference, Utrecht

-    (1990) Timed comprehension of binding in advanced L2 learners of English
    Language Learning 40:557-599

-    (1991) The poverty-of-the-stimulus argument and multicompetence. Second
    Language Research 7:103-117

-    (1995) Multicompetence and the effects of age. In Singleton, D. and Lengyel, Z. The
    Age Factor in Second Language Acquisition. Clevedon: Multilingual Matters

Coppieters, R (1987) Competence differences between native and near-native speakers
    Language 63:544-573

Crain S and Fodor, J.D. (1985)

-    - (1987) Sentence matching and overgeneration. Cognition 26:123-169

-    and Thornton, R. (1990) Paper presented at the University of Massachusetts workshop
    on the acquisition of Wh-movement

Cresti, D.(1995) Extraction and reconstruction.. Natural Language Semantics 3,1:79-122

Culicover, P (1993) Evidence against ECP accounts of the *that*-t effect. Linguistic Inquiry
    24.3:557-561

-    and Novak, A.(1995) Paradoxes and puzzles of triggering. Unpublished ms.

Curtiss, S.R. (1977) Genie: A Linguistic Study of a Modern Day "Wild Child". New York:
    Academic Press

- (1980)The critical period and feral children. UCLA Working Papers in Cognitive
    Linguistics,2:21-36

Davies, W.D. and Kaplan,T.I. (1998) Native speaker vs L2 learner grammaticality
    judgements. Applied Linguistics 19,2: 183-203

De Bot, K and Clyne, M. (1989) Language reversion revisited. Studies in Second Language
    Acquisition 11:167-177

Demonet, J.F., Wise, R. and Frackowiak, R.S. (1993) Language functions explored in nomal

subjects by positron emission tomography: a critical review. Human Brain Mapping 1:39-47

De Villiers, J., Roeper,T. And Vainikka, A.(1990) The acquisition of long-distance rules. In Frazier, L. and De Villiers, J. (eds) Language Processing and Acquisition. Dordrecht: Kluwer

Dutto, M.G. (1986) (ed) The Italians in Scotland. Edinburgh: Italian Consulate General

Ekstrand (1976) Age and length of residence as variables related to the adjustment of migrant children with special reference to second language learning. In Nickel, G. (ed) Proceedings of the Fourth International Congress of Applied Linguistics, vol.3: 179-198

Ellis, R (1991) Grammaticality judgements and second language acquisition. Studies in Second Language Acquisition 13:161-186

Elman, J.L. (1993) Learning and development in neural networks; the importance of starting small. Cognition 48:71-99

Emery, O. (1986) Linguistic decrement in normal aging. Language and Communication 6: 47-64.

Emmorey, K., Bellugi, U., Friederici, A. and Horn, P.(1995) Effects of age of acquisition on grammatical sensitivity: evidence from on-line and off-line tasks. Applied Psycholinguistics 16:1-23

Erteschik-Shir, N. (1973) On the nature of island constraints. Indiana University Linguistics Club

Ervin, S. and Osgood, C. (1954) Second language learning and bilingualism. Journal of Abnormal and Social Psychology 49(suppl.):139-146

Ervin-Tripp, (1974) Is second language learning like the first? TESOL Quarterly 9:111-127

Fathman, A (1975) The relationship between age and second language productive ability. Language Learning 25: 245-253

Feier, C. D. and Gerstmann, L. J. (1980) Sentence comprehension abilities throughout the adult life span. Journal of Gerontology 35:722-728

Flege, J.E. (1987) A critical period for learning to pronounce foreign languages? Applied Linguistics 8:162-177

Forster, K.I. and Stevenson, B.J. (1987) Sentence matching and well-formedness. Cognition 26:171-186

Frazier, L. and Fodor, J.D. (1978) The sausage machine: a new two-stage parsing model. Cognition, 6:291-325

Friedman, S. A . and Forster, K.I. (1985) The psychological status of overgenerated sentences. Cognition 19:101-131

Fucci, D., Ellis. L. and Petrosino, L. (1990) Speech clarity/intelligibility: test-retest reliability of magnitude estimation scaling. Perceptual and Motor Skills 70:232-4

Fukui, N. (1993) Parameters and Optionality in Linguistic Inquiry 24:399-420

Garnham, A. (1985) Psycholinguistics: Central Topics. London: Routledge

Gass, S. (1994) The reliability of second-language grammaticality judgements. In Tarone, E., Gass, S. and Cohen Research Methodology in Second Language Acquisition Hillsdale, N.J.:Lawrence Erlbaum Associates

Gathercole, S.E. (1998) The development of memory. Journal of Child Psychology and Psychiatry and Allied Disciplines. 39,1:3-27

-      and Baddeley, A.D. (1993) Working Memory and Language. Hove: Lawrence Erlbaum

Genesee, F. (1988) Neuropsychology and second language acquisition in Beebe L. M. Issues in SLA: Multiple Perspectives. New York: Newbury House

Goodluck, H., Sedivy, J. and Foley, M. (1989) Wh-questions and extraction from temporal

adjuncts: a case for movement. <u>Papers and Reports on Child Language Development</u> 28:123-130

- and Rochemont, M. (1992) <u>Island Constraints: Theory, Acquisition and Processing.</u> Dordrecht: Kluwer

- Foley, M. and Sedivy, J. (1992) Adjunct islands and acquisition. In Goodluck, H. and Rochemont, M. (eds) <u>Island Constraints: Theory, Acquisition and Processing</u> Dordrecht: Kluwer

- and Finney, M. (1993) When are chains constructed? in Schafer, A. (ed) <u>Proceedings of the 23rd North Eastern Linguistics Society Meeting</u> Graduate Linguistic Student Association, University of Massachusetts, Amherst

- Saah, K.K. and Stojanovic, D. (1995) On the default mechanism for interrogative binding. <u>Canadian Journal of Linguistics</u> 40,4: 377-404

Goldin-Meadow, S. (1982) The resilience of recursion: a study of a communication system developed without a conventional language model. In Wanner , ? and Gleitman, L.R. (eds) <u>Language Acquisition: State of the Art.</u> Cambridge: Cambridge University Press

Goldowsky, B.N. and Newport, E.I. (1993) Modeling the effects of processing limitations on the acquisition of morphology: the less is more hypothesis. In Clark,C. (ed) <u>The Proceedings of the 24th Annual Child Language Research Forum</u> 124-138. Stanford, CA: Center for the Study of Language and Information

Goss, N., Ying-Hua, Z. and Lantolf, J.P. (1994) Two heads may be better than one: mental activity in second language grammaticality judgments. In Tarone. E., Gass, S. and Cohen, S. <u>Research Methodology in Second Language Acquisition.</u> Hillsdale, N.J: Lawrence Erlbaum Associates

Grimshaw, G.M., Adelstein, A, Bryden, M.P., and MacKinnon, G.E. (1998) First-language acquisition in adolescence: evidence for a critical period for verbal language development. <u>Brain and Language</u> 63: 237-255

Grimshaw, J. and Rosen, S.T. (1990) Knowledge and obedience: the developmental status of the Binding Theory. <u>Linguistic Inquiry</u> 21: 187-222

- and Samek-Lodovici, V. (1998) Optimal subjects and subject universals. In Barbosa, Fox, P.D.,Hangstrom, P.M. McGinnis, M.and Pesetsky,D (eds.) <u>Is the Best Good Enough? Optimality and Competition in Syntax.</u> Cambridge, Mass.: MIT Press.

Grosjean, F (1982) <u>Life With Two Languages</u> Harvard University Press, Cambridge, Mass

Gunter, T.C., Jackson, T.L. and Mulder,G (1995) Language memory and aging - an electrophysiological exploration of the N400 during reading of memory-demanding sentences. <u>Psychophysiology</u> 32: 215-229

Haegeman, L. (1994) <u>Introduction to Government and Binding Theory</u> 2nd ed. Oxford:Basil Blackwell.

Haig, J. (1991) Universal grammar and second language acquisition: the influence of task type on late learners' access to the Subjacency principle. TESL monograph, McGill University

Harrington, M (1992) Working memory capacity as a constraint on L2 development. In R.J.Harris (ed) <u>Cognitive Processing in Bilinguals.</u> Amsterdam: Elsevier Science Publishers

Hawkins, J.A. (1994) <u>A Performance Theory of Order and Constituency.</u> Cambridge: Cambridge University Press

Hedgecock, J.(1993) Well-formed vs. ill-formed strings in L2 metalingual tasks: specifying features of grammaticality judgements. <u>Second Language Research</u> 9:1-21

Huang, C-Y. J. (1982) Logical relations in Chinese and the theory of grammar. Doctoral

Dissertation, MIT.

Hurford, J. R(1991) The evolution of the critical period for language acquisition. <u>Cognition</u> 40: 159-201

- (to appear) Functional innateness: explaining the critical period for language acquisition. In Noonan, M., Moravscik, E. and Newmeyer, F. <u>Functionalism and Formalism in Linguistics volume 2 (SLCS 42)</u>. Amsterdam: John Benjamins

Hyams N. (1987) The theory of parameter and syntactic development. In Roeper, T and Williams, E. (eds) <u>Parameter Setting</u> Dordrecht: D Reidel

Hyltenstam, K. (1992) Non-native featues of near-native speakers: on the ultimate attainment of childhood L2 learners. In Harris, R.J. (ed) <u>Cognitive Processing in Bilinguals.</u> Amsterdam: Elsevier Science Publishers

Ioup, G., Boustagui, E., El Tigi, M. and Moselle, M. (1994) Reexamining the critical period hypothesis: a case study of successful adult SLA in a naturalistic environment. <u>Studiesin Second Language Acquisition</u> 16:73-98

Isaki, E. and Plante E. (1997) Short-term and working memory differences in language/learning disabled and normal adults. <u>Journal of Communication Disorders</u> 30,6:427-437

Johnson, J. (1988) Critical period effects on universal properties of language: the status of Subjacency in the acquisition of second languages. Doctoral Dissertation, University of Illinois at Urbana-Champaign.

- and Newport, E. (1989) Critical period effects in second language learning: the influence of maturational state on the acquisition of English as a second language <u>Cognitive Psychology</u> 21:60-99

- (1991) Critical period effects on universal properties of language: the status of Subjacency in the acquisition of a second language <u>Cognition</u> 39:215-258

- (1992) Critical period effects in second language acquisition: the effect of written versus auditory materials on the assessment of grammatical competence. <u>Language Learning</u> 42,2: 217-248

    Shenkman, K.D., Newport, E.L. and Medin,D. (1996) Indeterminacy in the grammar of adult language learners. <u>Journal of Memory and Language</u> 35:335-352

Juffs, A.and Harrington, M. (1995). Parsing effects in L2 sentence processing: subject and object asymmetries in Wh-extraction. <u>Studies in Second Language Acquisition</u> 17:483-516

- - (1996) Garden path sentences and error data in second language sentence processing. Language Learning 46,2: 283-326

Jusczyk, P.W (1996) <u>The Discovery of Spoken Language</u>. Cambridge, Mass: MIT Press

Kausler, D.H. (1994) Learning and Memory in Normal Aging. San Diego: Academic Press

Kellerman, E.(1995) Age before beauty: Johnson and Newport revisited. In Eubank L., Selinker, L.,and Sharwood Smith, M. (eds) <u>The Current State of Interlanguage: Studies in Honor of William E. Rutherford</u>. Amsterdam: Benjamins

Kemper, S. (1987) Life-span changes in syntactic complexity. <u>Journals of Gerontology</u> 42: 232-238

- (1997) Metalinguistic judgments in normal aging and Alzheimer's Disease. <u>Journals of Gerontology: Psychological Series</u> 52B.3: P147-P155

- Kynette, D., Rash, S., O'Brien, K. and Sprott, R. (1989) Life-span changes to adults' language: effects of memory and genre. <u>Applied Psycholinguistics</u> 10:49-66

Kessler, C. and Idar, I. (1979) Acquisition of English by a Vietnamese mother and child. <u>Working Papers in Bilingualism</u> 18:65-79

Kim, K.H.S., Relkin, N.R., Lee, K-M and Hirsch, J (1997) Distinct cortical areas associated with native and second languages. <u>Nature</u> 388:171-174

King, J. and Just, M. (1991) Individual differences in syntactic processing: the role of working memory. Journal of Memory and Language 30: 580-602

Kluender, R. (1992) Deriving island constraints from principles of predication. In Goodluck, H. and Rochemont, M. (eds) Island Constraints: Theory, Acquisition and Processing Dordrecht: Kluwer Academic Press

- and Kutas,M. (1993) Bridging the gap: evidence from ERPs on the processing of unbounded dependencies. Journal of Cognitive Neuroscience 5,2: 196-214

Krashen, S. D. (1973) Lateralization, language learning and the critical period: some new Evidence. Language Learning 23:63-74

(1982) Principles and Practice in Second Language Acquisition Oxford: Pergamon Kynette, D., Kemper, S., Norman,   ' and Cheung, H.(1990)

- and   Kemper, S. (1986) Aging and the loss of grammatical forms: a cross-sectional study of language performance. Language and Communication 6:65-72

Lamendella, J. T. (1977) General principles of neurofunctional organisation and their manifestations in primary and nonprimary language acquisition. Language Learning 27:155-196

Lardiere, D. ((1998) Case and tense in the "fossilized" steady state. Second Language Research 14,1:1-26

Lenneberg, E. H. (1967) Biological Foundations of Language. New York: Wiley

Leow, R.P. (1997) Attention, awareness and foreign language behavior. Language Learning 47,3:467-505

Levelt W. J. M .(1972) Some psychological aspects of linguistic data. Linguistische Berichte 17:18-30

- Haans, J.A.W.M. and Meijers, A.J.A. (1977) Grammaticality, paraphrase and imagery. In Greenbaum, S. Acceptability In Language. The Hague: Mouton

Lightfoot, D. and Weinberg, A. (1988) Review Article: "Barriers" by Noam Chomsky Language 64:366-383

Long, M. (1990) Maturational constraints on language development. Studies in Second Language Acquisition 12:251-285

Lust, B. (1988) Universal Grammar in second language acquisition:promises and problems in critically relating theory and empirical studies. In Flynn, S .and O'Neil, W. Linguistic Theory in Second Language Acquisition Dordrecht: Kluwer

McKinnon, R. and Osterhout, L. (1996) Constraints on movement phenomena in sentence processing: evidence from event-related brain potentials. Language and Cognitive Processes 11,5:495-523

McLaughlin, B. (1984) Second Language Acquisition in Childhood Vol.1 (2nd ed.) Hillsdale, N. J.:Lawrence Erlbaum

Mack, M. (1984) Early bilinguals: how monolingual are they? In Paradis, M. and Lebrun, Y. (eds) Early Bilingualism and Child Development. Lisse: Swets and Zeitlinger.

Martohardjono, G. (1992) Wh-movement in the acquistion of English as a second language. Paper presented at the Workshop on the Acquisition of Wh-movement, University of Massachusetts, Amherst.

- and Gair,J.W. (1993) Apparent UG inaccessibility in second language acquisition: misapplied principles or principled misapplications? In Confluence: Linguistics, L2 Acquisition and Speech Pathology. Eckman, F. (ed) Amsterdam:Benjamins

- and Flynn, S. (1995) Is there an age factor for Universal Grammar?   In Singleton, D. and Lengyel, Z. The Age Factor in Second Language Acquisition Clevedon, Avon: Multilingual Matters

Mayberry, R.L. and Eichen, E.B. (1991) the long-lasting advantage of learning sign language in childhood: another look at the critical period for language acquisition. Journal of

Memory and Language 30:486-512

Miyake, A., Carpenter, P.A. and Just, M.A. (1994) A capacity approach to syntactic comprehension disorders – making normal adults perform like aphasic patients. Cognitive Neuropsychology 11,6:671-717

Murofeshi, K (1997) Numerical matching behavior by a chimpanzee (Pan troglodytes): subitizing and analogue magnitude estimation. Japanese Psychological Research 39,3:140-153

Murphy, V. A.(1997) The effect of modality on a grammaticality judgement task. ??????????

Nagata, H. (1992) Anchoring effects in judging grammaticality of sentences. Perceptual and Motor Skills 75:159-164

-    (1997) An assimilation effect in judging the grammaticalty of sentences violating the Subjacency condition. Perceptual and Motor Skills 84:755-767

Neisser, U. (1984) Interpreting Harry Bahrick's discovery: what confers immunity against forgetting? Journal of Experimental Psychology: General 113:32-35

Neufeld, G. G .(1979) Towards a theory of language learning ability. Language Learning 29:227-

Neville, H.J., Nicol, J.L., Barss, A. Forster, K.I. and Garrett, M.F. (1991) Syntactically based sentence processing classes: evidence from event-related brain potentials. Journal of Cognitive Neuroscience 3:151-165

Newmeyer, F. J. (1983) Grammatical Theory: Its Limits and Its Possibilities. Chicago: The University of Chicago Press

Newport, E. and Supalla, T. (1990) A possible critical period effect in the first language. Unpublished.ms, Psychology Department, University of Rochester

Novoa, L. Fein, D. and Obler, L. (1988) Talent in foreign languages: a case study. In Obler, L. and Fein, D. (eds) The Exceptional Brian: The Neuropsychology of Talent and Special Abilities. New York: Guilford Press

Nunnally, J.C. (1978) Psychometric Theory (2nd ed.) New York: McGraw Hill

Obler, L.K. and Mahecha, N.R. (1991) First language loss in bilingual and polyglot aphasics. In Seliger, H.W. and Vago, R.M. (eds) First Language Attrition .Cambridge: Cambridge University Press

Ojemann, G.A.(1991) Cortical organization of language. Journal of Neuroscience 11:2281-2287

Otsu, Y. (1981) Toward a theory of syntactic development. Doctoral dissertation. MIT.

Oyama, S. C. (1976) A sensitive period for the acquisition of a phonological system. Journal of Psycholinguistic Research 5:261-283

-    (1978) The sensitive period and comprehension of speech. Working Papers on Bilingualism 16: 1-17

Papagno, C., Valentine, T. and Baddeley, A.D. (1991) Phonological short-term memory and foreign language vocabulary learning. Journal of Memory and Language 30:331-347

-    and Vallar, G. (1995). Verbal short-term memory and vocabulary learning in polyglots. Quarterly Journal of Experimental Psychology Section A-Human Experimental Psychology 48,1:98-107

Patkowski, M. (1980) The sensitive period for the acquisition of syntax in a second language.Language Learning 30:449-72

-    (1990) Age and accent in a second language: a reply to James Emil Flege. Applied Linguistics, 11: 73-89

Paradis, M. (1977) Bilingualism and aphasia. In Whitaker, H.A. and Whitaker, H. (eds) Studies in Neurolinguistics (vol.3). New York: Academic Press

Penfield, W. and Roberts, L. (1959) Speech and Brain Mechanisms. New York: Atheneum

Phinney, M.(1987) The pro-drop parameter in second language acquisition. In Roeper, T

and Williams, E. Parameter Setting Dordrecht: Reidel

Pollock, J-Y. (1989) Verb movement, Universal Grammar and the structure of IP. Linguistic Inquiry 20:365-424

Pritchett, B.L. (1992) Grammatical Competence and Parsing Performance. Chicago: Chicago University Press

Pye, C., Cheung, H. and Kemper, S. (1992) Island constraints at eighty. Goodluck, H. and Rochemont, M. (eds) Island Constraints. Dordrecht: Kluwer

Rizzi, L. (1982) Issues in Italian Syntax Dordrecht: Foris.

- (1990) Relativized Minimality . Cambridge, Mass.: The MIT Press

- (1994) Argument/adjunct (a)symmetries. In Cinque, G., Koster, J., Pollock, J-Y., Rizzi, L and Zanuttini, R. (eds) Paths Towards Universal Grammar: Studies in Honor of Richard S Kayne. Washington D.C.: Georgetown U.P.

Robinson, P. (1997) Individual differences and the fundamental similarity between implicit and explicit second language learning. Language Learning 47,1:45-99

Roeper, T. and de Villiers, J. (1990) Ordered decisions in the acquisition of wh-questions. In Weissenborn, J., Goodluck, H. and Roeper, T. (eds) Theoretical Issues in Language Acquisition: Papers from the Boston Conference. Hillsdale, N.J.:Lawrence Erlbaum

Ross, J.R.(1967) Constraints on variables on syntax. Doctoral dissertation, MIT, published as Infinite Syntax! (1986) Norwood, N.J.: Ablex

Saah, K.K. and Goodluck, H. (1995) Island effects in parsing and grammar; evidence from Akan. The Linguistic Review 12: 381-409

Schachter, J. (1988) Second language acquisition and its relationship to Universal Grammar Applied Linguistics 9:221-235

- (1989) Testing a proposed universal. In Gass, S. and Schachter, J. Linguistic Perspectives on Second Language Acquisition.Cambridge, Mass: Cambridge University Press

- and Yip, V. (1990) Grammaticality judgements: why does anyone object to subject extraction? Studies in Second Language Acquisition 12: 379-392

- (1990) On the issue of incompleteness in second language acquisition. Second Language Research 6:93-124

Schmidt, R. W. and McCreary, C.F. (1977) Standard and super-standard English: Recognition and use of prescriptive rules by native and non-native speakers. TESOL Quarterly 11:415-429

- (1981) Interaction, acculturation and the acquisition of communicative competence. University of Hawaii Working Papers in Linguistics 13/3:29-77

- (1990) The role of consciousness in second language learning. Applied Linguistics 11:129-158

- (1993) Awareness and second language acquisition. Annual Review of Applied Linguistics13:206-226

Schneiderman, E.I. and Desmarais, C. (1988) The talented language learner: some preliminary findings. Second Language Research 4: 91-109

Schumann, J. (1978) The Pidginization Process: A Model for Second Language Acquisition Rowley, Mass: Newbury House

- and Stauble, A-M. (1983) A discussion of second language acquisition as decreolization. In Andersen, R.(ed) Pidginization and Creolization as Language Acquisition. Rowley, Mass.: Newbury House

Schwartz, B.D. and Gubala-Ryzak, M.(1992) Learnabiltiy and grammar reorganization in L2A:against negative evidence causing the unlearning of verb movement. Second Language Research 8,1:1-38

- and Sprouse, R.(1996) L2 cognitive states and the Full Transfer/Full Access model.

Second Language Research 12.1:40-72

Scovel T. (1977) The perception of foreign accents by children, adults and aphasics and its relationship to the notion of a critical period for language acquisition. In Dressler, W.U. Proceedings of the XII International Congress of Linguists

-       (1981) The recognition of foreign accents in English and its implications for psycholinguistic theories of language acquisition. In Savard, J.-G. (ed) Proceedings of the 5th Congress of AILA. Laval, Canada: University of Laval press

Selinker, L. (1972) Interlanguage. International Review of Applied Linguistics X:209-30

Service, E. (1992) Phonology, working memory and foreign-language learning. Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology 45,1: 21-50

Singleton, D. (1995) Introduction: a critical look and the critical period hypothesis in second language acquisition research. In Singleton, D. and Lengyel, Z. (eds) The Age Factor in Second Language Acquisition. Clevedon: Multilingual Matters

Smith, N. and Tsimpli, I. (1991) Linguistic modularity? A case study of a 'savant' linguist. Lingua 84: 315-351

Snow, C. (1987) Relevance of the notion of a critical period for language acquisition. In Bernstein, M. (ed) Sensitive Periods in Development: An Interdisciplinary Perspective. Hillsdale, N.J.: Erlbaum

-       and Meijers, G. (1977) On the secondary nature of syntactic intuitions. In Greenbaum, S. (ed.) Acceptability In Language. The Hague: Mouton

-       and Hoefnagel-Hohle, M. (1978) The critical period for language acquisition. Child Development 49:1114-1128

Sorace, A (1990) Indeterminacy in first and second languages: theoretical and Methodological issues. In De Jong, J.A.H. and Stevenson, D. (eds) Individualizing the Assessment of Language Abilities. Clevedon: Multilingual Matters

-       (1993) Incomplete vs. divergent representations of unaccusativity in non-native grammars of Italian. Second Language Research 9:22-47

Spolsky, B. (1995) Measured Words. Oxford: Oxford University Press

Stowe, L. (1986) Parsing wh-constructions:evidence for on-line location. Language and Cognitive Processes 1:227-248

Tarone, E. (1988) Variation in Interlanguage. London: Edward Arnold

Tavakolian, S. (1981) The conjoined clause analysis of relative clauses. In Tavakolian, S. (ed) Language Acquisition and Linguistic Theory Cambridge, Mass: MIT Press

Thogmartin, C. (1982) Age, individual differences in musical and verbal aptitude and pronunciation achievement by elementary school children learning a foreign language. International Review of Applied Linguistics 20:66-72

Tomaselli A. and Schwartz, B.D. (1990) Analysing the acquisition stage of negation in L2 German: support for UG in adult SLA. Second Language Research 6:1-38

Toner, M. and Emmanuel, F. (1989) Direct magnitude estimation and equal appearing interval scaling of vowel roughness. Journal of Speech and Hearing Research 32:78-82

Travis, L. (1988) Linguistic theory, neurolinguistics and second language acquisition. In Flynn,S.and O'Neil. W. (eds) Linguistic Theory in Second Language Acquisition Dordrecht: Kluwer

Trudgill, P. (1983) Sociolinguistics Harmondsworth: . Penguin.

Tsimpli, I-M. and Roussou, A. (1991) Parameter resetting in L2? UCL Working Papers in Linguistics 3

-       and Smith, N. (1991) Second language learning: evidence from a polyglot savant. UCL Working Papers in Linguistics 3

Uziel, S. (1993) Resetting universal grammar parameters: evidence from second language acquisition of Subjacency and the Empty Category Principle. Second Language Research 9.1:49-83

Vainikka, A and Young-Scholten, M (1996) Gradual development of L2 phrase structure. In Second Language Research 12,1:7-39

Vargha-Khadem, F., Carr, L.J., Isaacs, E., Brett, E., Adams, C., and Mishkin, M.(1997) Onset of speech after left hemispherectomy in a 9 year-old boy. Brain 120,1: 159-182

Walsh, T.M. and Diller, K.C. (1986) Neurolinguistic considerations on the optimum age for second language learning. In Diller, K.C. (ed) Individual Differences and Universals in Language Learning Aptitude. Rowley, Mass: Newbury House

Weber-Fox, C.M. and Neville, H.J. (1996) Maturational constraints on functional specializations for language processing: ERP and behavioral evidence in bilingual speakers. Journal of Cognitive Neuroscience 8,3:231-256

Weinreich, U. (1953;1968) Languages in Contact: Findings and Problems. The Hague: Mouton

Wexler, K. and Culicover, P (1980) Formal Principles of Language Acquisition. Cambridge, Mass: The MIT Press

- and Manzini, R. (1987) Parameters and learnability in Binding Theory. In Roeper, T. and Williams, E. Parameter Setting. Dordrecht: Reidel

White, L. (1988) Island effects in second language acquisition. In Flynn, S. and O'Neil, W. (eds) Linguistic Theory in Second Language Acquisition. Dordrecht: Kluwer

- (1989) Universal Grammar and Second Language Acquisition Amsterdam: John Benjamin

- (1990) The verb-movement parameter in second language acquisition. Language Acquisition 1:337-360

- (1991) Adverb placement in second language acquisition: some effects of positive and negative evidence in the classroom. Second Language Research 7:133-161

- (1992) Universal Grammar and second language acquisition: Current Trends and New Directions to appear in Ritchie, W. and Bhatia, T. (eds) Handbook of Language Acquisition Academic Press

- and Genesee, F. (1996) How native is near-native? The issue of age and ultimate attainment in adult second language acquisition. Second Language Research 12:233-265

- and Juffs, A. (1998) Constraints on wh-movement in two different contexts of non-native language acquisition: competence and processing. In Flynn, S, Martohardjono, G and O'Neil, W. (eds) The Generative Study of Second Language Acquisition Hillsdale, N.J.: Lawrence Erlbaum Associates:

Wode, H. (1994) Nature, nurture and age in language acquisition: the case of speech perception. Studies in Second Language Acquisition 16:325-345

Young-Scholten, M. (1994) On positive evidence and ultimate attainment in L2 phonology. Second Language Research 10,3:193-214

Zatorre, R.J. (1989) On the representation of multiple languages in the brain: old problems and new directions. Brain and Language 36:127-147

Zobl, H. (1990) Evidence for parameter-sensitive acquisition: a contribution to the domain-specific vs. central processes debate. Second Language Research 6:39-59

- (1992) Sources of linguistic knowledge and uniformity of nonnative performance. Studies in Second Language Acquisition 14:387-402

Zurif, E., Swinney, D., Prather, P., Wingfield, A. and Brownell, H. (1995) The allocation of memory resources during sentence comprehension: evidence from the elderly. Journal of Psycholinguistic Research 24,3: 165-182

ITALIAN SPEAKERS OF ENGLISH

MAIN STUDY 1997-98

INSTRUCTIONS - SCRIPT

Rationale: I'm trying to find out if the age someone arrived in this country from Italy has an effect on their knowledge of English grammar: if there's a difference between people who arrived in their early 20s, for example, and people who were in their 30s when they came here.

Introduction: You're going to read 86 English sentences and decide whether you think they're OK or not. Some will be OK, some will be rubbish, and some will be in between - nearly OK perhaps or not quite complete rubbish. You'll be giving each sentence a number, according to what you think of it. As this is quite a strange thing to do, there are two practices, before you start the actual task.

Practice 1: this is with lines.
[Read instructions]
Reminder: you don't have to try to be mathematically accurate - just work on your first impressions. The first line will be roughly in the middle of the range of lines. Your first number can be anything - 21/2, 98, 1005, but you must be able to go higher and lower - and remember that you can't use zero or minus numbers. Remember that the first line will remain in view throughout.

After Practice 1: was that OK? [elicit problems etc]

Practice 2: this is with short sentences.
[Read instruction]
Reminder: This time, the higher the number you give the better the sentence sounds to you. Again, the first sentence is roughly in the middle - its not the very best, and its not the worst. Don't spend too long thinking about each sentence - just go by your first impressions. Try not to look back at your earlier numbers, and don't worry about forgetting what numbers you've already given. Just keep moving ahead. The first sentence will remain in view throughout.

After Practice 2: was that OK? [elicit problems etc]

Task
[Read instructions]

Any questions?
You'll only have a short time to read each sentence

Instructions for task

Now you'll be shown another series of English sentences. As with practice 2, some of the sentences are OK - possible English sentences - some are terrible, and completely impossible; some are in between - perhaps they sound nearly right, or not quite impossible. You'll be reading the sentences, but try to imagine they're spoken. Once again, read the first sentence, and write down any positive number for it . Don't spend time thinking about the sentence - we're interested in YOUR FIRST IMPRESSIONS. Read the next sentence and write down another number **according to how possible you think it is compared to the first sentence**. For example, if you think it sounds twice as good as the first sentence, you would give it twice the first number; if you think it's half as possible, you'd give it half the first number, and so on. Please try not to think in terms of fixed upper and lower limits like 1-10 or 1-100.

YOU WILL HAVE ONLY A SHORT TIME TO SEE EACH SENTENCE
THE FIRST SENTENCE WILL REMAIN IN VIEW

Appendix I:2 Questionnaire for Italian Speakers of English

97exper/interview/nns
Interviewee no: _____

PERSONAL INFORMATION

It would be of great help to the investigation if you could complete the questionnaire below. All information given will be kept strictly confidential.

1. Age group at time of test:

20-25 ☐   26-30 ☐   31-35 ☐   36-40 ☐
41-45 ☐   46-50 ☐   51-55 ☐   56-60 ☐
61-65 ☐   66-70 ☐

2. Which part of Italy (eg Region, city) were you born in? please ✓

1. Valle D'Aosta
2. Piemonte
3. Lombardia
4. Veneto
5. Venezia Giulia
6. Trentino/Alto Adige
7. Emilia Romagna
8. Liguria
9. Toscana
10. Umbria –
11. Le Marche
12. Lazio
13. Abruzzi
14. Molise
15. Campania
16. Puglia
17. Basilicata
18. Calabria
19. Sardegna
20. Sicilia
21. Torino
22. Milano
23. Venezia
24. Firenze
25. Roma
26. Napoli

If you were born outside Italy please ✓   and say where............

3. How old were you when you completed your full-time education?
....

4. How old were you when you first arrived in Britain to live?............

5. Why did you or your family move to Britain?
please ✓

marriage ☐   work ☐   study ☐   other ☐

6. Did you learn English before you moved to Britain? Yes ☐   No ☐

7. If you answered "yes" to question 6, please indicate how you learned English before moving to Britain, and approximately for how long:

1 School ....years
2 University ....years
3 Private Language school ....years
4 Other ....years

8. Did you attend English classes after your arrival in Britain? If so, for how long (approximately)?

Yes ☐   for.............. months/years

No ☐

9. Which language do you mainly speak at home now?

English ☐   Italian ☐   Both ☐   Other ☐

10. If you would like to add any other information about how, where and when you use your two languages, please do:

....
....
....
....

Thank you very much for your help

Appendix II   Test materials

1 Adjunct extraction

1.1 control: 1.1.1   I don't know where George gave his lecture after he left the pub yesterday
1.1.2   Which problem does Dr Smith believe his students can solve?
1.1.3   Didn't Ann want to go home after drinking a bottle of vodka?
1.1.4   I wonder where Fred will want to move to after selling his flat.

1.2 subjacency violation (extraction from adjunct):
1.2.1   Which pub did George give his lecture after he left yesterday?
1.2.2   What does Dr Smith believe his students can solve the problem because of
1.2.3   What do you wonder where Fred will want to move to after selling?
1.2.4   What did Ann want to go home after drinking a bottle of with her friends?

1.3 ECP violation (extraction of adjunct from adjunct)
1.4.1   Where did George leave the pub after going yesterday?
1.4.2   How long does Dr Smith believe which problem his students can solve after working on it?
1.4.3   Where did Ann want to go home after drinking a bottle of vodka with her friends?
1.4.4   for how much do you wonder which part of the city Fred will choose to move to after selling his flat?

2 Relative clause extraction

2.1 control: 2.1.1   I bought a book my friend had written a chapter about hillwalking in.
2.1.2   My daughter phoned the police about the jewellery she thought someone had stolen
2.1.3   They wondered whether their uncle had sent the parcel of books they collected
2.1.4   Can you understand why the college chose to organise a course the students dislike?

2.2 subjacency violation:
2.2.1 What did you buy the book in which my friend has written a chapter about?
2.2.2 What did my daughter phone the police on the day when someone stole?
2.2.3 What did they wonder whether it was their uncle who had sent a parcel of?
2.2.4 What did you believe the students who disliked ?

2.3 ECP violation
2.3.1 Who have you bought the book in which had written a chapter about?
2.5.2 Who did their daughter phone the police about the jewellery which stole?
2.5.3 From where did their daughter do anything about the jewellery which was stolen

3 Extraction from Noun Complement

3.1 Control   3.1.1. Have you heard the news that Jo has got the job?
3.1.2 They spread the rumour that the headteacher had punished the boys.

3.1.3 We don't believe the story that the princess toured the island

3.1.4 Has Andrew heard the news that the firm is moving to London?

3.2. Subjacency violation (object extraction)

3.2.1 What did you hear the news that Jo has got?

3.2.2 Who did they spread the rumour that the headteacher had punished?
   (adjunct extraction)

3.2.3 Which castle did you believe the story that the princess toured the island after visiting?

3.3 ECP violation (subject extraction)

3.3.1 Who have you heard the news that has got the job?

3.3.2 Who did they spread the rumour that had punished the boys?

3.3.3 Who did you believe the story I told you last night that toured the island?
   (adjunct extraction)

3.3.4 Where did you believe the story that the princess toured?

3.3.5 When has Andrew heard the news yesterday that the firm is moving to London?

4 Extraction from wh-islands

4.1 Control 4.1.1 I wonder which story about Tom's travels was the children's favourite.

4.1.2 Do the doctors know whether the medicine cured the disease?

4.1.3 Do you know whether John said he moved to Ireland at the age of 10?

4.1.4 I can't imagine why Mary chose not to speak to the Director

4.2.Subjacency violation parameterised- (finite embedded clause)

4.2.1 I bought a book which I couldn't imagine why my friend had written a chapter in.

4.2.2 Their daughter phoned the police about the jewellery which she wondered whether someone had stolen.

4.2.3 Did they collect the parcel which they wondered whether their uncle had sent?

4.2.4 The students dislike the course which I can't understand why the college organised.
   (nonfinite embedded clause)

4.2.5 I've read the book which I can understand why they decided to publish.

4.2.6 They dislike the course which I wonder why the college chose to organise.

4.2.7 Who do imagine why Mary chose not to speak to?

4.3 Subjacency violation (object extraction)

4.3.1 Which story do you wonder whether the children enjoyed?

4.3.2 Who do the doctors know whether the medicine cured?

4.3.3 What do you know why John told about moving to Ireland?
   (subject extraction)

4.3.4 What do you wonder which story about was the children's favourite?

4.3.5 What do the doctors know whether the discovery of cured the disease?

4.3.6 Where do you know whether the man from said he moved to Ireland at the age of 10?

4.3.7 Who do you imagine why the parents of chose not to speak to the Director?

4.4 ECP violation (subject extraction)

        4.4.1 What do you wonder whether was the children's favourite?

        4.4.2 What do the doctors know whether cured the disease?

        4.4.3 Who do yo imagine why chose not to speak to the Director?

        4.4.4 Who do you know whether they said, in the hotel lounge, moved to Ireland?
           (adjunct extraction)

        4.4.5 Where do you know whether John said he moved to at the age of 10?

        4.4.6 At what age do you know whether John said he moved to Ireland?

        4.4.7 Where did they wonder whether their uncle had sent the parcel?

        4.4.8 Where do the students dislike the course which I can't understand why the
           college organised

Grammatical fillers:

1. I imagine George went home after leaving the pub yesterday
2. Does Dr Smith believe his students can solve the problem?
3. I'm sure Ann wanted to go to bed after drinking a bottle of vodka with her friends
4. I imagine Fred will want to mover to the country after selling his flat.
5. My friend wrote a chapter about hillwalking in the book I've just bought.
6. Their daughter phoned the police about her stolen jewellery yesterday.
7. They thought their uncle had sent them a parcel of books.
8. The college decided to organise a new course but the students dislike it.
9. I've just heard that Jo has got the job.
10. The headteacher had punished the boys before the parents were told.
11. They said the princess toured the island by car.
12. Has Andrew heard that the firm is moving to London for economic reasons.
13. The story about Tom's travels was the children's favourite.
14. The doctors are sure that the medicine cured the disease.

Nonsense fillers:

1. Went after he yesterday I George home imagine pub he left.
2. His Dr solve does can Smith students problem think the?
3. Bottle to heard to wanted bed vodka Ann I a drinking go after of.
4. Different to will I Fred selling buy imagine want somewhere house after flat his a.
5. In my hillwalking friend new about wrote book a chapter my.
6. The police yesterday jewellery daughter their phoned he stolen about.
7. Books a thought their of sent they parcel uncle them had.
8. Rumour punished the before had headteacher the started boys the.