Cognitive Patch Theory: A Comparison of the Morphosyntactic Competences of Advanced ESL Learners and Native Speakers of English

By

Amer M.Th. Ahmed

A thesis submitted in conformity with the requirements for the degree of Master of Arts
Department of Curriculum, Teaching and Learning
Ontario Institute for Studies in Education of the University of Toronto

© Copyright by Amer M.Th. Ahmed (2011)
Abstract

This study investigates the morphosyntactic competence of advanced ESL learners and native speakers of English. Using the framework of the Government and Binding approach (Chomsky, 1981, 1986), the study tests the predictions made by the evolved Fundamental Difference Hypothesis (Bley-Vroman, 2009), namely that the grammars of advanced L2 learners are unreliable (where reliability means converging to the L2 grammar), non-convergent to the L2 grammar, and characteristic of patches (where patches are extragrammatical principles independent of the normal syntactic processes). The participants of the study were tested on three tasks (timed grammaticality judgment task, a correction task, and a preference task). The findings of the study indicate that the difference between the morphosyntactic competence of the advanced ESL learners and that of native speakers is gradient rather than categorical.
Acknowledgments

I would like to thank Dr. Rena Helms-Park, my thesis supervisor, for providing me with very helpful feedback, which made this thesis much better both in content and in the general layout.

I am also indebted to Dr. Gabriela Alboiu at the Department of Languages, Literatures and Linguistics of York University for kindly discussing with me all the test sentences administered in this study and providing me with very helpful guidance.

My thanks also go to Dr. Nina Spada, my second reader, for reading the final version of this thesis and providing me with very helpful input.

A word of thanks should also go to all the participants of the study for their participation, and to the teachers and coordinator at the Faculty of Continuing Studies of the University of Toronto for their collaboration.
# Table of Contents

## Chapter 1: Introduction
1.1 Introduction ........................................................................................................ 1  
1.2 Research Questions .......................................................................................... 5

## Chapter 2: Review of the Literature ................................................................. 9
2.1 Introduction ...................................................................................................... 9  
2.2 Generative Approaches to L2A ..................................................................... 9  
2.2.1 The Full Transfer Full Access Hypothesis (FTFA) ....................................... 9  
2.2.2 The Full Access No Transfer Hypothesis ...................................................... 13  
2.2.3 The Partial Access Hypothesis ................................................................. 14  
2.2.4 The Fundamental Difference Hypothesis (FDH) ......................................... 16  
2.3 Cognitive Approaches to L2A ....................................................................... 18  
2.4 Theoretical and Empirical Approaches to L2A ................................................. 20  
2.5 The Original Fundamental Difference Hypothesis ....................................... 26  
2.6 The Evolved Fundamental Difference Hypothesis: Cognitive Patch Theory .... 30

## Chapter 3: The Method ....................................................................................... 36
3.1 Participants ..................................................................................................... 36  
3.2 Materials ....................................................................................................... 37  
3.2.1 Grammatical Categories ............................................................................. 41  
3.2.2 Extragrammatical Categories: Linguistic Patches ....................................... 53  
3.3 Procedures ..................................................................................................... 57

## Chapter 4: Data Analysis .................................................................................... 61
4.1 Testing the Reliability and Convergence Hypotheses in Task 1 ....................... 61  
4.2 Testing the Reliability and Convergence Hypotheses in Task 2 ....................... 63  
4.3 Testing the Reliability and Convergence Hypotheses in Task 3  
   (the Correction Task) ......................................................................................... 65  
4.4 The ESL Group: Testing the Patching Hypothesis ......................................... 66

## Chapter 5: Discussion and Conclusion .............................................................. 68
5.1 Discussion ...................................................................................................... 68  
5.2 Conclusions ................................................................................................. 70  
5.2.1 Pedagogical Implications .......................................................................... 71  
5.3 Limitations of the Study and Suggestions for Future Research ....................... 72

## References .......................................................................................................... 75

## Appendix A ........................................................................................................ 79

## Appendix B ........................................................................................................ 96
List of Tables

Chapter 2: Review of the Literature
Table 1. The difference between native language development and foreign language learning (Bley-Vroman, 2009) 32

Chapter 3: The Method
Table 1. A biographical information about the participants 36
Table 2. Breakdown of test sentences 39

Chapter 4: Data Analysis
Table 1. The overall performance of the ESL group and NS group on task 1 (TGJT) 62
Table 2. The overall performance of the ESL group and NS group on task 2 64
Table 3. The overall performance of the ESL group and NS group on task 3 65

Appendix A:
Table 1. The syntactic categories and test sentences 79
Table 2. The linguistic patches 89
Table 3. The overall performance of individual ESL participants and NSs on task 1 (TGJT) 91
Table 4. The overall performance of individual ESL participants and NSs on task 2 (the correction task) 93
Table 5. The overall performance of the ESL group and the NS group on task 3 (the preference task) 93
Table 6. The performance of the ESL group on linguistic patches 95
<table>
<thead>
<tr>
<th>List of Appendices</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Appendix A</td>
<td>79</td>
</tr>
<tr>
<td>Appendix B</td>
<td>96</td>
</tr>
</tbody>
</table>
Chapter 1:

Introduction

1.1 Introduction

This study aims to see whether the grammars of advanced English as a second language (ESL) learners are reliable or not, convergent to that of the second language (L2) grammar or not, and/or characteristic of the normal syntactic processes or of other extragrammatical processes. The theoretical framework is the Government and Binding approach (Chomsky, 1981, 1986). The choice of this framework rather than the new version (i.e. the Minimalist Program, Chomsky, 1995) is motivated by the fact that the Minimalist Program is not “a hypothesis about language or a new approach to language, displacing earlier ones” and the fact that the Minimalist Program “is theory neutral”. In fact, “whatever one’s conception of UG, one can be interested in principled explanation (MP), or not. And if so, essentially the same questions will arise” (Chomsky, 2008, p. 157). The study falls within one of the generative approaches to second language acquisition (SLA), that is, the Fundamental Difference Hypothesis, particularly its most recent formulation (Bley-Vroman, 2009).

The inception of the cognitive revolution in linguistics in the mid fifties of the last century as realized by the Chomskyan generative enterprise has changed the history of the study of language forever. Within this theory, language is not viewed merely as a system of habits that children learn from their surroundings through conditioning, a dogma that

---

1 Throughout this study, the term SLA will be used to refer to the field of second language acquisition, whereas the term L2A will refer to the process itself.
dominated the field of language study unchallenged during the heyday of behaviorism in psychology. As Chomsky argued in his famous critique (1959) of Skinner’s behaviorism (1957), a child does not come to this world as a *tabula rasa* that is fully shaped by experience (i.e. environmental conditions). Far from that, children come to the world equipped with a language faculty whose initial state is called Universal Grammar (UG), a genetically inbuilt system of principles that constrains the form any human language can take. It is assumed to be guiding the child learning its first language by drastically limiting the range of possible hypotheses that a child can entertain. Some of the principles (i.e. submodules) of UG are the Subjacency Principle, the Binding Principle, the Bounding Principle, Theta Theory, Case Theory, Empty Category Principle (ECP). These principles are assumed to be fixed in that children universally have access to them and make use of them to learn their first language.

In addition to these, other UG principles are not universally fixed, namely parameters. These allow variation, and are therefore responsible for the variation among languages. Yet, these parameters are also assumed to be restricted in the amount of variation they allow, and, in most cases are assumed to have one or the other of two values [+- feature] (Chomsky 1981, 1986).

The history of the relationship between second language acquisition (L2A) and UG seems to go back as far as the early seventies of the last century, when studies began to look at L2A from a generativist’s approach (Thomas, 2004, p. 184). Since then, there has been a continuous debate on whether or not UG is available in L2A in the same way as it is in first language acquisition (L1A), and if yes, to what extent is UG available in L2A. The research
has also investigated the role that first language (L1) plays in L2A and whether or not the L1 parameters are transferred into the L2A or whether all the L1 parameters are reset to the L2 values (White, 2003, pp. 15-6).

Looking at the current literature, one can identify three main approaches to the relationship between L2A and UG (as formulated by White 2003; Kanno, 1998, for example). The first main approach is the Full Access Hypothesis. Two distinct strands can be traced within this framework. The first is what is called the Full Access No Transfer Hypothesis, represented by Kanno (1998), who believes that L2 learners have full access to the UG principles, and that no transfer from L1 takes place in the process of L2A. The second version within the Full Access is the Full Access Full Transfer Hypothesis (FAFT) (Schwartz and Sprouse, 1994, 1996; White, 1985; Belikova and White, 2009; Slabakova, 1999). According to this version of Full Access, L2 learners start with transferring L1 parameters in the L2A, and only later are able to reset their parameters to the values of L2 system through exposure to the L2 input.

The second main approach to L2A is that of the Partial Access Hypothesis. There are three versions of this approach. Tsimpli and Rousou (1991) claim that functional categories\(^2\) are no more accessible in adult L2 learning, since these are subject to maturation after the critical period of language learning (Lenneberg, 1967). The second strand maintains that some features of the functional categories are no more accessible (Hawkins, 2007, Hawkins

\(^2\) The functional category module is implicated in the acquisition of grammatical categories such as number, gender, definiteness, case, tense and aspect, mood, agreement, etc. There are a number of functional projections, namely IP/TP, AgrSP, AgrOP, MP. In the Minimalist Program, these are vP, ProgP, PerfP, NegP, MP, TP (Tallerman, 2005, pp. 53-61; Adger, 2003).
The third main approach to L2A and UG is called the Fundamental Difference Hypothesis (FDH), represented by Schachter (1988), Bley-Vroman (1990) and Bley-Vroman and Yashinaga (2002). These scholars maintain that the underlying mechanisms of L1A and those of L2A are essentially different, and that an SLA theory, independent of UG, is needed. However, taking recent developments in linguistics, especially minimalism (Chomsky, 1995), psychology, mathematical learning theory, Bley-Vroman (2009) has introduced the cognitive patch theory as an alternative for SLA research, and it is this theory that will be adopted in the present study. According to Bley-Vroman (2009), the idea of patching comes in the aftermath of the demise of a rich and highly articulated UG with minimalism. To account for the obvious differences between L1A and adult L2A, Bley-Vroman (2009) postulates that patches or extragrammatical principles are peripheral in L1A but predominant in adult L2A. This ties in with research that is being done in neurolinguistics, where two distinct memory systems are postulated to account for the difference between L1A and L2A. According to this approach, the difference between L1A and L2A is marked by a shift from procedural memory system heavily implicated in the processing of grammatical computations to the declarative/lexical memory system, thought to be responsible for lexical learning (Ulman, 2001). Consistent with this research program, theoretical and empirical approaches to instructed SLA also seem to be heading in the same direction. Ellis (2005) proposes two distinct research tools that aim to tap into L2 learners’ implicit (procedural) and
explicit (declarative) knowledge systems, and suggests that different tasks enhance these two distinct knowledge systems. Other SLA researchers call attention to the role of selective attention in adult L2A and hold the view that no learning is possible without “awareness” (Schmidt, 1990). Furthermore, they propose that while implicit learning may be possible, explicit learning through instruction is also facilitative (Schmidt, 1990). Still others call for using insights from generative linguistic theory to inform pedagogical practice (Rothman, 2010; Rutherford & Sharwood Smith, 1985). This takes the form of enhancing the input L2 learners receive, for example, by calling their attention to the way parameters are set in the L2, thus helping to make certain implicit knowledge about the L2 explicit through instruction.

1.2 Research Questions

The purpose of this study is to contribute to the field of SLA by attempting to answer the following questions:

1. Do the interlanguage grammars of advanced L2 learners of English in the area of morphosyntax converge to one another? In other words, do the interlanguage grammars resemble one another or not?

2. Are the interlanguage grammars of advanced L2 learners of English in the area of morphosyntax reliable or unreliable? In the framework of this study, and following Bley-Vroman (2009), reliability means that the acquired grammars should converge to the L2 grammar.
Are the interlanguage grammars of advanced L2ers of English in the area of morphosyntax characteristic of patches or not? In the framework of the present study, and following Bley-Vroman (2009), patches are extragrammatical principles independent of the normal syntactic processes.

The results gained from this study have the potential to provide insights into the effects of instructional intervention on the process of L2A. To be specific, if the study were to find that the interlanguage grammars of advanced L2 learners of English are unreliable, and characteristic of patches, then a natural pedagogical implication would be that adult L2 learners can not rely on their own resources to acquire an L2 successfully. In other words, an unreliable and parasitic learning mechanism (in the sense that is unsuccessfully imitates the L2 learning mechanism) would not be trusted under any theory of SLA be it generative or otherwise to function properly on its own. This would provide a solid argument for instructed SLA. Furthermore, if such is the case, another implication would follow, namely that the field of instructed SLA is heading in the right direction by focusing on the findings of studies conducted within the frameworks of Processing Instruction (PI), the Interaction Hypothesis, and Focus-On-Form (FOF). In other words, these instructional studies would help by making the unreliable, non-convergent to the L2 grammar, patchy language acquisition device work more reliable through manipulating the input that adult L2 learners receive. This will hopefully make certain feedback trigger the kind of ‘awareness’ that is needed in order for the language acquisition system to become more efficient. This can be
done by giving more weight to negative evidence (both explicit and implicit) as an instructional intervention.

If, on the other hand, the results of the study showed that the interlanguages grammars of advanced L2 learners of English are reliable (i.e. convergent to the L2 grammar, and not “patchy”), then a more prudent pedagogical approach would be to give more weight to positive evidence and less to negative evidence. This is because the interlanguage grammar would then be up to the challenge and trustworthy, and learners would mostly need opportunities to be exposed to rich L2 input, something that some approaches to UG (i.e. some of the Full Transfer Full Access Hypothesis proponents) and some linguists (e.g. Krashen, 1981, 1985) have been calling for.

To sum up, the present study is an attempt to provide insights to both L2A theory and L2 pedagogy. This will be done by testing one of the most recent theories of L2A, namely the cognitive patch theory (Bley-Vroman, 2009). To this end, Chapter 2 will provide a review of generative and cognitive approaches to L2A, including the history of the hypothesis to be tested, both in its original form (the Fundamental Difference Hypothesis, Bley-Vroman, 1990) and its most recent evolution (the Cognitive Patch Theory, Bley-Vroman, 2009). Chapter 3 will set the stage for the data analysis by providing a description of the participants who took part in the study, the method of analyzing the data of the study, the syntactic categories targeted in the study, and the procedure to be followed for screening participants and collecting data. Chapter 4 will be dedicated to analyzing the data and providing the results in relation to each of the hypotheses to be tested. Chapter 5 will
conclude the study with a discussion of the implications of the study for L2A theory, future research, and pedagogy. This chapter will also include a section devoted to the scope and limitations of the study and suggestions for future research.
Chapter 2

Review of the Literature

2.1 Introduction

This chapter will summarize the essentials of the three major approaches, that is, the generative approaches to L2A, the cognitive approaches to L2A, and the theoretical and empirical approaches to L2A. The Fundamental Difference Hypothesis (FDH) (Bley-Vroman, 1990) and the evolved Fundamental Difference Hypothesis (cognitive patch theory of L2A, Bley-Vroman, 2009), fall within the generative approaches to L2A but also merit their own category (as discussed in the previous chapter), since they are the focus of the present study.

2.2 Generative Approaches to L2A

2.2.1 The Full Transfer Full Access Hypothesis (FTFA).

One of the generative approaches to L2A is the Full Transfer Full Access hypothesis (FTFA), introduced by Schwartz and Sprouse (1994, 1996). According to this theory, L2 learners start by fully transferring everything except specific phonological properties from L1 to L2. In other words, they start learning an L2 with a natural grammar from their L1, and only later will they reset the parameters according to those of the L2. This means that there will be no parameter resetting in those properties where there is not enough positive evidence (i.e. input) from L2, as the divergence between L1 and L2 is only due to the lack of enough positive evidence from L2. An example of the FTFA is the placement of adverbs in the performance of French learners of English. When French speakers start learning English,
they encounter sentences such as ‘she often drinks coffee’, but this may not be enough to tell them that those clauses with adverbs between the verb and the object are incorrect in English (*she drinks often coffee), since they are grammatical in French. The result is that they will continue to allow both representations, and their English performance in this area will consequently diverge from that of native speakers of English. Slabakova (1999) investigated the UG Aspect Parameter in the performance of three different proficiency levels of Bulgarians learning English. The results of the study showed that the low intermediate learners were transferring the L1 parameter of aspect to their L2, whereas high intermediate and advanced students were not.

The study also showed that resetting the parameter to the L2 value co-occurred with acquiring those related grammatical structures, as predicted by the FTFA hypothesis. In another study of the Pro[noun]-Drop Parameter and a number of hypothetically related grammatical properties such as missing subjects, free subject verb inversion, and that trace effects, White (1985) administered a grammaticality judgment test to 54 Spanish students learning English with five different levels of proficiency, and had as a control group 19 French learners of English with five different levels of proficiency. The fact that Spanish is a (+ pro – drop) language, whereas English and French are (– pro- drop) languages would mean that the Spanish learners of English would initially think of the following English sentences as grammatical, whereas the French group would not.
The results of the study showed that the Spanish students carried over the L1 parameter into their L2 performance whereas the French were not. The study also found that the Spanish students’ transfer of their L1 parameter was less with the higher-level proficiency in the L2 English. However, losing an L1 parameter did not result in an accompanying loss of all the associated grammatical properties as predicted by the FTFA hypothesis, since the results were mixed here.

Following recent developments in linguistic theory (Chomsky, 1995, Uriagereka, 1999), Belikova & White (2009) review the findings of various studies addressing the subjacency condition (Chomsky, 1973), which has been used a touchstone test to either support access to UG by L2 learners (e.g. White & Juffs, 1998, as cited in Belikova & White, 2009) or as evidence against access to UG by L2 learners (e.g. Bley-Vroman, Felix, & Ioup, 1988; Schachter, 1989, 1999, as cited in Belikova & White, 2009). Two important insights come out of this study. First, all the studies reviewed show that L2 learners were much better at rejecting extractions of wh-words out of strong islands, i.e. subjects, adjuncts (including relative clauses) than rejecting extractions of wh-words out of weak islands, i.e. wh-islands and N(oun) complements. To illustrate this point, consider the following examples:

10 a. [Which city], did [you witness [the destruction of]]?

b. * [Which city], did [you visit Paris [before]]?

1. a.* Which book did John hear [a rumour [that you had read]]?

---

1 The subjacency condition is a UG constraint, which states that “a constituent may not move over more than one bounding category at a time”, where the bounding categories are “[a] sentence (S) and N(oun)P(hrase) (now inflection phrase [IP] and determiner phrase [DP], respectively” (Belikova & White, 2009, p. 202).

2 The term ‘island’ is used in the generative literature to refer to those cases where extraction of a wh-word out of certain structural constituents is impossible. It was first introduced by John Ross (1967).
b. * Which book did John meet [a child [who read ,]]?

As can be seen from the above examples, while the general pattern in the reviewed studies was that native speakers of English rejected all of the above, many of them accepted sentences such as (10a) above, as it is an example of a weak island, i.e. extracting a wh-word out of a noun complement ‘of _’. As for L2 learners, they performed much better on strong islands than on weak islands (10b, 11a, 11b), i.e. they rejected a lot of strong islands, while accepting a lot of weak islands (10a). This, according to the authors, provides evidence that L2 learners are sensitive to the ambivalent status of weak islands (Belikova & White, 2009, pp. 209-211).

The second important point that comes out of this study is the authors’ claim that developments in linguistic theory (Chomsky, 1995, Uriagereka, 1999) have made it impossible to tease out L1 effects from UG effects. In particular, Uriagereka (1999) has made it clear that “all languages, including languages without wh-movement, have subjects and adjuncts, [and that] the general mechanism that renders subjects and adjuncts inaccessible to further internal manipulation, including extraction, will therefore be universal”.

Herschensohn (, 2000, 2009) believes that the differences between L1 and adult L2 language learning processes are quantitative rather than qualitative, and that “the most blatant A[ge]o[f]A[quisition] effects are phonetic inaccuracy, morphological errors, indeterminate G[rammaticality]J[udgments], and processing speed, factors that essentially do not qualify as poverty-of-the-stimulus effect [barring GJs which]] indicate a short-coming in that respect”. However, even indeterminate GJs “are quantitatively, not qualitatively, different from those
of native speakers” as was evidenced by the performance of “many learners who were indistinguishable from native speakers in GJs, regardless of AoA” (pp. 269, 282).

In general, the proponents of the FTFA hypothesis believe that UG guides both L1A and L2A; therefore, instructed L2A does not have any effect on the L2 acquisition process (Doughty, 2003, p. 257). Schwartz acknowledges a role for negative evidence; yet, she “questions” the role such evidence play in “engaging” UG (Gass, 2003, pp. 226-227). However, unlike Schwartz and Sprouse (1993, 1996, as cited in Doughty, 2003, p. 257), White (1987, 1991, as cited in Doughty, 2003, p. 257) is of the view that instructed L2A is necessary sometimes, especially when the L2 is a proper subset of the L1 with respect to a certain parameter, as L2 learners “will have to retreat from the overly general hypothesis that emanates from their L1[…], something which cannot be done on the basis of positive evidence alone”. Herschensohn (2000) also believes that both positive evidence (i.e. input) and negative evidence (i.e. corrective feedback) are required in L2 acquisition.

2.2.2 The Full Access No Transfer Hypothesis

Another approach within the Full Access Hypothesis is the Full Access No Transfer strand. Kanno (1998) found that the English L2 learners of Japanese were successful at observing two UG principles. L2 learners were found to differ from the native speakers in that at least at the early stages of L2 learning, their success at one UG principle is not a predictor of their success at other UG principles, and that the L2 learners’ access to the UG principles are not consistent, unlike the case with native speakers whose observance of the UG principles is consistent (pp. 386-387).
2.2.3 The Partial Access Hypothesis

A further generative approach to L2A is the Partial Access Hypothesis. Hawkins (2007) refers to three types of evidence that L2 learners have some specifically linguistic innate properties that are UG driven and that constrain their L2 behavior. The first evidence comes from a UG principle called the Overt Pronoun Constraint (OPC), which, states that: “Where there is an alteration in a language between an overt/null pronoun, only the null pronoun can take a quantified expression as antecedent (e.g. someone, nobody, which student)” (p. 466).

Consistent with the above-stated principle, Pérez-Leroux and Glass (1997, as cited in Hawkins, 2007, pp. 467-468) showed that both English L2 learners of Spanish and native speakers performed similarly in that both of them used a null pronoun when this was the preferred form and an overt pronoun when it was the favored form. The results according to Hawkins (2007) cannot therefore be attributed to the negative input the learners receive because such negative evidence is not available. Nor can the results be attributed to L1 transfer because English always uses the overt pronoun in contexts similar to the ones given in the above-mentioned experiments. Moreover, there is no evidence that such an alternation in the use of the pronoun is taught in the classroom. Hawkins therefore concludes than only an assumption of a UG driven constraint can account for the participants’ behavior. Tsimply and Rousou (1991) and Smith and Tsimpli (1995, as cited in Kong, 2005, p. 230) argue that L2 learners do have access to some grammatical options, which are neither part of their L1 grammatical system nor that of their L2 grammatical system. Yet, these grammatical options
are natural as they are constrained by UG. However, following Borer (1983, as cited in Kong, 2005, p. 230) and Chomsky (1991, as cited in Kong, 2005, p. 230) who assume that parameters are independent from UG principles and are mainly associated with lexical items particularly functional categories (e.g. determiners, inflections, etc.), Tsimpli and Rousou (1991) and Smith and Tsimpli (1995, as cited in Kong, 2005, p. 230) assume that functional categories are subject to maturation during childhood, hence, “parameter-resetting in L2 is excluded” (Tsimpli & Roussou, 1991, p. 151). Of special interest for the purpose of this study is the proposal that “general learning mechanisms” are responsible for the adoption of the correct parametric value at the more advanced levels in adult L2A (p. 152). Hawkins and Chan (1997, as cited in Kong, 2005, p. 231), while agreeing with Tsimpli and Rousou (1991, as cited in Kong, 2005, p. 230-231) that languages may differ parametrically, hold the opinion that it is not the grammatical categories that are inaccessible to L2 learners but rather certain features\(^3\) of such grammatical categories as ‘agreement’ and ‘determiner’. Thus, post Critical Period L2 learners will not have access to features of certain grammatical categories, but will nonetheless have access to the grammatical categories themselves. This approach is therefore called the Failed Functional Features hypothesis. According to this approach, two predictions are possible in L2A. The first is that L2 learners will map the L2 morphological features on to those of L1. The second prediction is that continued exposure to L2 will help the learners move progressively from their L1 to their L2. However, givingen that certain grammatical features are inaccessible to L2 learners, the learners will develop certain

\(^3\)A grammatical feature is a property that is always associated with certain grammatical categories. Examples of such features are gender, person, number, tense etc. English, for example has the tense feature (+/- past) and the person feature (1st, 2nd, 3rd person) (see White, 2003, p. 10; Tallerman, 2005, p. 232)
grammatical features which are neither those of L1 nor those of L2 but they are nonetheless *natural* as they are constrained by UG principles. Beck (1998, as cited in Kong, 2005, p. 231) found that even advanced English learners of German did not recognize certain grammatical features in German (i.e. verb raising in German) and concluded that certain grammatical features are not accessible to older L2 learners. As a result, the author proposed another approach, called the Local Impairment Hypothesis.

2.2.4 *The Fundamental Difference Hypothesis (FDH)*

In stark contrast to the above positions is an approach called the Fundamental Difference Hypothesis (FDH). The proponents of this approach (Schachter, 1988; Bley-Vroman, 1990) believe that the underlying mechanisms involved in the learning of L1 and those in L2 are essentially different (see, however, McDonald, 2000 for a different account of the observed difference). Therefore, L2A researchers should seek a non-UG theory of L2A.

A proponent of Lenneberg’s Critical Period Hypothesis (CPH) (1967), Schachter (1988, pp. 222-223) provides four arguments to support her position that L1 and L2 acquisition processes are very different. The first of these arguments is that of completeness. No L2 learner, no matter how proficient they are, can acquire a mental system (i.e. a grammatical competence with its subcomponents of phonology, morphology, syntax, and semantics), which is the same as that acquired by every normal native speaker (Coppieters (1987, as cited by Schachter, 1988, pp. 223-224). The second argument is equipotentiality defined as “the ability to learn any natural language X in the same amount of time and with

---

4For an overview of the L2A research studies that attempt to resolve the CLH debate, see Hyltenstam & Abrahamsson, 2000).
the same ease as the time and effort required to learn a completely unrelated language Y” (p. 225). Children are equipotential to learn any natural language in the same amount of time and with the same ease, whereas adult L2 learners are not equipotential to any natural language (Ard and Homburg (1983, as cited in Schachter, 1988, p. 226). The third argument is that any study of L2 learners’ performance would show an L1 transfer (see Gass and Selinker, 1983; Kellerman, 1984, as cited by Schachter, 1988, p. 226). And the final argument is the common instances of fossilization\(^5\) that appears in adult L2 language performance but not in L1 child production. Therefore, the author claims that any theory of L2A should look much more seriously at the issue of fossilization (Schachter, 1988, p. 229).

Bley-Vroman and Yoshinago (2000) studied native and non-native speakers’ judgments of the grammaticality of sentences with multiple wh-phrases, as in the following:

15. Who is eating what?
16. Who is sitting where?
17. Who is singing where?
18. Who is walking when?
19. Who is getting down how?
20. Who is crying why? (Bley-Vroman and Yoshinaga, 2000, p. 10)

The results of the study demonstrated that while native speakers of English rated sentences with ‘what’ and ‘where’ (i.e. sentences 18 and 19) as grammatical and the rest were rated with decreasing degrees of acceptability, the Japanese L2 learners of English only rated

\(^5\) Fossilization “[is] the regular reappearance or re-emergence in IL [interlanguage] productive performance of linguistic structures which were thought to be eradicated” (Selinker, 1975, as cited in Schachter, 1988, p. 228)
sentence (18) as being grammatical, and the rest were all rated as more or less unacceptable. The authors concluded that the results could not be attributed to L1 transfer in any possible manner, as all of the above structures are completely grammatical in Japanese. Instead, they argue based on the results of their study for the importance of L2 input and noticing, both of which have to be available for some of the L2 learners to approach a competence similar in general to that of native speakers. In the context of this study, this means that learners who have not encountered and noticed L2 multiple wh-phrases will not be able to incorporate them in their L2 competence system. (Bley-Vroman and Yashinaga, pp. 21-24)

2.3 Cognitive Approaches to L2A

Ulman (2001) posits two distinct knowledge systems, namely declarative and procedural knowledge systems with their neural correlates, namely declarative memory and procedural memory respectively. Declarative memory system “has been implicated in the learning, representation, and use of knowledge about facts ‘semantic knowledge’ and events ‘epistemic knowledge’”. It is a system that is not “informationally encapsulated”, thus accessible to “multiple response systems” and is available for conscious recollection. On the other hand, procedural memory system “has been implicated in the learning of new, and control of long-established, motor and cognitive ‘skills’ or ‘habits’ (e.g. from simple motor acts to riding a bicycle to skilled game playing).” It is thus a system which is “informationally encapsulated”, as “neither the learning nor the remembering of these procedures appears to be accessible to conscious memory-thus the system is often referred to as an “implicit memory” system” (pp. 106-107). This model postulates two distinct
cognitive systems for the lexicon and the grammar with “at least partially distinct neural correlates”. The model also posits that the lexicon system is an associative memory system (i.e. learned through association, analogy, frequent exposure). For example, using this system, speakers can form the plural to the novel noun *nog* by analogizing to the familiar one *dog*, whereas the grammar system “underlies symbol-manipulation” This system is the main source of productivity in language since it operates over variables. For example, a variable (e.g., Noun stem) can represent a large number of examples (e.g. dog, cat, door). It is this system that allows speakers to generalize linguistic knowledge to any novel item. Thus, using the algorithm (*N*plural = *N*stem + s), speakers can form plural nouns to any novel item forming nouns such as *dogs, cats, doors*, etc. The model moreover postulates different neural localizations for the two components, namely the lexicon and the grammar (pp. 107-108). Ulman extends his declarative/procedural model of L1 to that of L2 by positing that later exposure to a language, especially during late childhood or post puberty leads to more reliance on declarative/lexical memory. Thus, while in L1, grammatical forms are compositionally computed in the procedural memory, they are memorized in L2 in the declarative/lexical memory. However, age is not the only factor, as extensive practice of L2 should “increase the language relative dependence on procedural memory for grammatical computations” (but see Doughty, 2003, pp. 291-298 for holding the opposite view). This should not; however, mean that age is not a factor after all. It should instead be interpreted as assigning roles to both age and practice in the implication of procedural memory in L2 (pp. 108-110). Montrul (2009), reviewing three studies (Hakansson, 1995; Montrul et al., 2008a,
2008b, as cited in Montrul, 2009) also found evidence for the hypothesis that adult L2 learners have grammatical forms and to a lesser degree morphological forms stored in declarative/lexical memory, whereas early bilinguals tested on their knowledge of L1, which they hadn’t used since childhood, showed evidence of employing their procedural memory when performing on oral tasks, which required the use of automatized and implicit knowledge. The study also lends some support to the FDH, as early bilinguals were (at least at the individual level) found to pattern with native speakers (NSs) rather than with non-native speakers (NNSs).

### 2.4 Theoretical and Empirical Approaches to L2A

Ellis (2005, p. 143) notes that “declarative knowledge is explicit and encyclopedic … [whereas] procedural knowledge is highly automated.” Also, implicit knowledge “displays lower variability than explicit knowledge [and whereas] implicit knowledge involves automatic processing; explicit knowledge entails controlled processing.” Most important to the purpose of the present study is the “cautious” claim that “explicit knowledge is learnable at any age, whereas implicit knowledge is not” (Ellis, pp. 149-150). Of particular importance to the present study is the fact that in his study, Ellis (pp. 157-158) found empirical evidence to support the prediction that the use of timed grammaticality judgments (TGJs) is a valid measurement of the learners’ implicit knowledge (see, however, Juff & Harrington, 1995, as cited in McDonald, 2000, p. 398 for a different view). Thus, commenting on these tasks in his study, he points out that the TGJ “task encouraged the use of feel, it was time-pressed, and there was little need or opportunity to access metalinguistic knowledge, [whereas] the
untimed [UGJ] encouraged a high degree of awareness and was unpressured.” These results are also consistent with Montrul’s (2009) finding that different tasks tap into different knowledge systems. In particular, she found that adult L2 learners were not different from early bilinguals in tasks that encouraged the use of explicit/declarative knowledge such as UGJs and written tasks, whereas in others such as oral production tasks, early bilinguals outperformed adult L2 learners. Despite that, DeKeyser (2003, p. 320) argues that “while speeded tests undoubtedly are more problematic for the retrieval of explicit than implicit knowledge …, time pressure does not guarantee a pure measure of implicit knowledge”, but then he adds that “researchers have to content themselves with eliciting knowledge under conditions that are more or less conducive to the retrieval of implicit and explicit knowledge, and then infer to what extent learning may have been implicit or explicit”. Other L2A researchers put forward the hypothesis that no learning is possible without noticing, where noticing is defined as the selective attention of individual learners to specific aspects of the form of the L2. According to this view, adults, unlike children, selectively pay attention to different linguistic stimuli in the input, and it is only when they start paying attention to specific forms that input turns into intake (i.e. acquired knowledge) (Schmidt, 1990). In this regard, Schmidt cites some experimental evidence, which seems to show that “the major change from child to adult consciousness is a shift from a passive mode that includes an awareness of the environment to more controlled mode that includes the strategic allocation of attention” (p. 145).
Of particular interest is the fact that all work that has been done to understand consciousness has defined unconscious processes as manifestations of “specialized systems operating in parallel [i.e. interacting], are not limited by short term capacity” nor are they “under voluntary control” and they are “difficult to modify”. They are therefore “fast, efficient, and accurate, and are responsible for skilled performance and most detailed cognitive processing”. On the other hand, conscious processes have always been considered a “manifestation of a limited capacity central processor” and they “are partially subject to deliberate control” Therefore, they are “slow, inefficient, mostly serial, and effortful” (Schmidt, p. 138). The connection to the implicit/explicit knowledge and to procedural/declarative memory systems is obvious. Thus, keeping in mind the fact that adults seem to use their attentional resources strategically and differ in the kinds of stimuli they pick from the linguistic input, it seems reasonable to assume a role for noticing and its relation to instruction, which would then help “making formal features of the target language more salient and facilitating input encoding” (p.149). Schmidt (1998, 2001, as cited in Gass, 2003, p.244) “has modified” his strong view that no learning is possible without noticing. He therefore acknowledges that learning can come about without the L2 learners being aware of that. However, he claims that such learning does not play a major part in L2A. In a similar vein, Dekeyser claims “there is little hard evidence of learning without awareness” (2003, p. 319).

The above literature on procedural/declarative memory systems and the implicit/explicit knowledge systems will be made use of in designing tasks that tap into both
memory and knowledge systems for the purpose of the present study. Also, the ‘Noticing Hypothesis’ has important pedagogical implications relevant to the present study. To be sure, the Noticing Hypothesis has laid the grounds for two promising instructed L2A research strands (and probably most of the pedagogical research): Processing Instruction (PI) research and Focus-On-Form (FOF) research. The overarching principle for both strands is the idea that while children’s processing mechanism of their L1 input is quite efficient since it is innately constrained by a limited research space, adult L2A learners are “disabled” in that their processing of the L2 input is attuned to that of their L1 and they do not seem to resort to an appropriate processing mechanism suitable to match the L2 input if they were “left to their own devices” (Doughty, 2003). Thus, instructed L2A couched in processing terms can assist in “organizing the processing space, hence perhaps re-enabling mechanisms that depend upon perceptual acuity” (p. 299). To overcome the problem of mismatch between the L1 processing strategy and the L2 input, IP instruction studies work at the level of the utterance and aim to direct the L2 learners’ attention to the fact that the L1 processing strategy is not appropriate to process the L2 input, and informs the learners of the L2 “cues” that they should be looking for instead (p. 288). According to this view, input is very important. In this respect this view is similar to the Interaction Hypothesis (Long, 1996, as cited in Gass, 2003, p. 229). Proponents of the Interaction Hypothesis acknowledge the role both of positive evidence (correct sentences given in the speech and writing of NSs or competent speakers). They also assign a privileged role to implicit negative evidence (i.e. corrective feedback).

6The Interaction Hypothesis accords a special importance to the role of conversation in L2A, and holds that conversation is not only a tool for learning but also an act of learning, as the interlocutors (i.e. a NS or a more competent interlocutor and a NNS negotiate for meaning, a process, which leads to the acquisition of knowledge of specific aspects of the L2 system by the NNS (Gass, 2003, pp. 234-235).
that is provided for NNSs in the form of recasts\(^7\) during conversation with NSs or more competent speakers of L2.

Of special importance to the purpose of the present study is the fact that such implicit negative evidence, while effective in the case of phonological and lexical errors, is only effective with surface low-level morphosyntactic errors (DeKeyser, 2003). One possibility why this is so is the fact that phonological and lexical errors affect meaning in conversation; therefore L2 learners notice them, and this can affect change in their interlanguage. Morphosyntactic errors, on the other hand, generally do not affect meaning during conversation, hence not attended to or noticed by L2 learners (Gass, 2003, p. 246). Gass (pp. 247-248) acknowledges the limitations of negative evidence provided through conversation, and suspects as a possibility that “surface-level phenomena can be learned, but abstractions cannot”. It is for this reason that DeKeyser (2003, p. 334) favors explicit learning over implicit learning, since for him “[t]he harder it is to learn something through simple association, because it is too abstract, too distant, too rare, too unreliable, or too hard to notice, the more important explicit learning processes become”. In fact, he claims that adults are better at learning using explicit knowledge, whereas children are better at learning using their implicit knowledge, and this is a function of the fact that there is a “shift during childhood from implicit to explicit processes” (cf. Schmidt, 1990; Ulman, 2001).

Focus on Form (FOF) instruction studies work “both within and across utterances”. The idea is that certain features of the L2 system, which the L2 learners need but fail to

\(^7\text{Recasts “refer to those instances in which one interlocutor rephrases an incorrect utterance with a corrected version, while maintaining the integrity of the original meaning” (Gass, 2003, p. 239).}
notice for some reason, require some pedagogical intervention (p. 289). DeKeyser (2003, p. 336) suggests that “focus on form is necessary to make learners consciously notice the abstract patterns that are not learnt implicitly”.

Of interest from a pedagogical perspective is also Rothman’s (2010) call for a more faithful incorporation of “linguistic description” in the pedagogical grammar textbooks, “as these serve the main source of reference to both teachers and adult learners alike”. To be sure, “all domains of grammar as presented in pedagogy can benefit from linguistic insights and will likely improve convergence toward the native target if done effectively” (p. 62). Others make a similar call for using insights from generative linguistic theory in pedagogical practice. Thus, consciousness-raising (C-R) in the classroom will be a continuum of explicit-implicit-nonexistent depending on the parameterized properties of L1 and L2. Using the PRO-Drop Parameter as an example, Spanish and Italian learners of English whose first language is [+ pro drop] would require an explicit instruction in the fact that English is a [- pro-drop], whereas English learners of Spanish or Italian would require an implicit instruction about the fact that these languages are [+ pro-drop]. (Rutherford & Sharwood Smith, 1985).
2.5 The Original Fundamental Difference Hypothesis

In his original formulation of the Fundamental Difference Hypothesis, Bley-Vroman (1990) identifies ten essential characteristics of adult foreign language learning\(^8\) language learning (p.6).

1. Lack of success: While normal children unanimously achieve the same level of competence in first language acquisition, the same is not true of adult L2 learners, who generally do not achieve a native-like competence in the second language. Unanimous success in first language acquisition and a general lack of success in adult second language learning thus stems from a theory of domain-specific (i.e. UG–driven) language acquisition in the former and a domain-generic problem-solving skills in the latter (Bley-Vroman, 1990, p.6).

2. General failure: While complete success is guaranteed for normal children in first language acquisition, complete success in adult L2A is “extremely rare, or perhaps even non-existent, especially as regards ‘accent’ and the ability to make subtle grammaticality judgments” (Bley-Vroman, 1990, pp. 6-7; Schachter, 1988). In fact, the rare cases of success can be disregarded in L2A in the same way that pathological cases of lack of success are disregarded in L1A (Bley-Vroman, 1990, p. 7).

3. Variation in success, course, and strategy: Among adult L2 learners, there is a “substantial variation” in the degree of attainment even when age, exposure, instruction, and other factors are held constant (Bley-Vroman, 1990, p.7). Similarly, there is also a variation

---

\(^8\)The terms ‘foreign’ and ‘second’ as used here to describe the learning of any language other than the child’s first language by adult learners are used interchangeably throughout.
in the learning path L2 learners follow, and the learning strategies they use in learning a
second language. This fact bears striking similarities to other general skill learning among
adults. In fact, it is exactly what one is to expect in adult general learning skills (pp. 7-8).

4. Variation in goals: Adult L2 learners differ not only in the degree of attainment,
but also in the “type of attainment” (Bley-Vroman, 1990, p. 8). Some learners have
developed good communicative skills with “rudimentary grammatical devices”. Others have
put more emphasis on grammatical correctness with less successful communicative skills.
Some develop communicative skills enough to help them function successfully in one
situation such as waiting on tables or lecturing in philosophy. Some are good at
pronunciation but less so in terms of grammar. This is consistent with a theory of learning of
general problem-solving skills, where goal setting is paramount. Children, on the other hand,
are only guided in learning their first language by a ‘goal’, which is genetically
predetermined and not under the learner’s control (p.9).

5. Correlation of age and proficiency: Studies which correlate age of acquisition and
ultimate attainment show that the younger the better. Interestingly, though, the fact that
teenagers in the age range of 10 and 15 are highly successful in achieving a native-speaker
competence suggests that Lenneberg’s (1967) conjecture that puberty is a cut-off point can
not be correct (Bley-Vroman, 1990, p.9).

6. Fossilization: While fossilization, a phenomenon frequently observed in adult L2
learners who seem to stabilize at a certain stage of learning “short of success” is the norm in
adult L2A, this phenomenon is never observed in children learning their first language, as the
latter groups’ brains remain plastic until success is achieved. The phenomenon bears striking resemblance to the fossilization that is often observed when one learns the techniques of sports such as the high jump in the wrong way, which would then make “defossilizing” extremely difficult for coaches (Bley-Vroman, 1990, pp. 9-10; Schachter, 1988).

7. Indeterminate intuitions: Even very advanced non-native speakers (NNSs) provide indeterminate performance on grammaticality judgment tests, which suggests that the cognitive system that underlies such performance cannot be the same cognitive system that underlies the performance of native speakers (NSs) on such tests (Bley-Vroman, 1990, pp. 10-11).

8. Importance of instruction: While there is no experimental evidence showing conclusively that instruction is necessary in adult L2A, some studies “seem to show” that instruction is helpful in adult L2A. Again, this suggests that adult L2A is very similar to the learning of general skills, where instruction and practice are indispensable. This stands in sharp contrast to the process of L1A, where the futility of such explicit instruction and practice is not even a matter of debate (Bley-Vroman, 1990, p.11).

9. Negative evidence: All serious attempts to construct formal first language theories (e.g. Wexler and Culicover, 1983, Pinker, 1990, as cited in Bley-Vroman, 1990, p.12) “assume that negative evidence is not used and that success is possible nonetheless”. This is not the case with adult L2A, where negative evidence “is at least sometimes useful, and sometimes, though not always, necessary”.

10. Role of affective factors: While the role of affective factors, such as personality, socialization, motivation, attitude, and so on, are “absolutely indisputable” in adult L2A, the same can not be true in L1A, where the process is “controlled by the development of an innate domain-specific faculty”. Once again, the phenomenon is reminiscent of general skill learning, which is “highly susceptible to such ‘affective factors’” (Bley-Vroman, 1990, pp. 12-13).

Despite the above-stated general characteristics of adult L2A, and the fact that there are “very few adults, if any [who] are completely successful”, there are “many who achieve very high levels of proficiency, given enough time, input, and effort, and given the right attitude, motivation, and learning environment”. According to Bley-Vroman, therefore, the logical problem of adult L2A is to “explain the quite high level of competence that is clearly possible in some cases, while also permitting a wide range of variation that is observed” (Bley-Vroman, 1990, p.13). And since different studies conducted within the UG framework have shown various UG effects in the performance of NNSs, something that cannot be attributed to chance, (e.g. Bley-Vroman, Felix & Ioup, 1988; Richi, 1987; White, 1985), the proposal made by Bley-Vroman would be to assume that the domain-specific faculty (i.e. UG) would in adult L2A be filled by the native language knowledge and a general abstract problem-solving system (p.13).

The adult L2 learners will use their knowledge of the native language to construct “a kind of surrogate for Universal Grammar”. However, this indirect knowledge of Universal
Grammar through the native language is “incomplete and accidental” and since it also depends on the individual learner’s ability to construct a UG-surrogate, one can expect some partial success, little chance of perfect success, and some considerable individual variation (see for example, Bley-Vroman et al. 1988 for results, which correspond to these predictions). Much of the observed ‘relative’ success in adult L2A can thus be attributed to the indirect knowledge of UG that the adult L2 learner brings to the process of L2A. Part of the burden of explaining adult L2A can also be shouldered on the adult L2 learner’s access to the immensely powerful problem-solving capacity, which cannot dedicate its full resources to the specific process of language learning. In mathematical linguistic terms, because the general human cognition has to deal with various systems, it might find it difficult to pick the language system from such a large set of systems, thus making the learning of a foreign language impossible without negative evidence. (p. 22).

2.6 The Evolved Fundamental Difference Hypothesis: Cognitive Patch Theory

Bely-Vroman (2009) concedes that the language faculty is not as completely domain-specific and independent from the other cognitive faculties as it was thought before. Instead, the language faculty interacts with other domain-generic faculties. In an assessment of the obvious differences between L1 development and L2 learning, and following Pullum and Scholz (2002), Bley Vroman suggests that they can be reduced to two principal contrasting properties, convergence and reliability. In Bley-Vroman’s words, convergence “is a relationship among the acquired grammars of learners exposed to the same target language”

---

The term ‘accidental’, which is used to refer to the indirect knowledge of UG by the adult L2 learner, has to do with the fact that some of the UG features, namely the setting of a certain parameter to the value of the L2 system might or might not be available in the adult learner’s native language (L1).
whereas reliability “is […] a relationship between the target grammar and the acquired grammar”. The two properties are related, but while reliability leads to convergence, the converse is not true; different unreliable acquisition systems might converge to the same nontarget grammars (e.g. grammars, which encoded specificity rather than definiteness, or in which the relative clauses had resumptive pronouns, etc.) (p. 178). Thus, in contrast to the learning theory of L1, where the learners equipped with an acquisition device, unanimously acquire the same target L1 given the L1 input, the L2A theory has to allow for the full range from the so-called near native cases to those cases where the acquired grammar is dramatically different from the target grammar. In other words, while the theory of L1A is predictive and deterministic, any L2A theory should be unpredictable and permissive (p. 178). The outcome of the L2A theory would be a set of “(not necessarily finite) interlanguage grammars”. This set will no doubt have an internal structure whose characteristics will be heavily contingent on the amount of input received and the structure of the native language, both essential factors that will influence the range of possibilities in the outcome grammar, in addition to a range of other non-linguistic factors (i.e. psychological and affective factors) (pp. 178-179). Notwithstanding the above differences, both native languages and interlanguages share the property of being “languages” in the sense that both involve a mapping between “human conceptual-intentional structures (i.e. meaning) and a physical expression realized using the speech apparatus [i.e. sounds]”, and in both “[e]lements can be combined with other elements to make larger units, and this can go on recursively”. In addition, foreign languages “ are processed in the human brain …and are
used to communicate, for private thinking, and for all functions of native languages” (p. 181). It is indeed the case as Bley-Vroman puts it succinctly that “[a]s rich UG falls away, the FDH, and indeed all UG-based research, must rethink its basic assumptions” (p. 183). One caveat here is that saying that the language faculty processes are not unique to language does not mean that there is no language faculty (p. 185). Rather, the language faculty should be conceived of as a system of processes, whose components can be shared by other systems. The important developments in the mainstream generative theory have rendered the debate between access or no access to UG meaningless, since the UG principles “minimal as they are-are instantiated in all languages” (Bley-Vroman, p. 186). Table 1 summarizes the difference between the native language development and foreign language learning.

Table 1. The difference between native language development and foreign language learning (Bley-Vroman, 2009).

<table>
<thead>
<tr>
<th>Native Language development</th>
<th>Foreign language learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Reliability: Children always succeed at language learning.</td>
<td>Foreign language learning is not reliable</td>
</tr>
<tr>
<td>2. Convergence: Children end up with systems that are so similar to those of others in the same speech community.</td>
<td>Foreign language learning is not convergent</td>
</tr>
<tr>
<td>3. Native languages are languages</td>
<td>3. Interlanguages are languages</td>
</tr>
</tbody>
</table>

To find a way out of this dilemma while at the same time accounting for the properties of nonconvergence and unreliability, the benchmarks of foreign language learning, Bley-Vroman (p. 187) suggests that “elements of the language faculty itself provide the seeds of nonconvergence and unreliability”. To be sure, linguists have noted that there are cases in
language where the core grammatical system doesn’t seem to be working. Consider the following examples (taken from Morgan, 1972; McCawley, 1988, as cited in Bley-Vroman, 2009, pp. 188-189)

21. ??Are/is (either) John or his parents here?
22. ??Are/is (either) John’s parents or his wife here?
23. He can’t seem to do it right (where can’t clearly doesn’t have scope over seem).
24. The usual men were drunk (where men are not usual).

Morgan suggested that

a patch is a way of employing the syntactic means provided by the language so as to allow one to say things of the same sort that can normally be expressed by a given construction in some class of cases that the normal syntactic means of the language do not provide for (Morgan, 1972, as cited in Bley-Vroman, 2009, p. 189)

So the “normal syntactic means” can be understood to mean the normal acquisition system (i.e. UG). In fact, the importance of such phenomena in language has become so obvious to mainstream generative linguists that Lasnik and Sobin (2000) have developed a theory called virus theory to account for such phenomena. According to the theory, viruses (which behave very much like patches) “are superficial and parasitic on the existing structures: These are shallow principles”. Sobin (1997) has actually noted such phenomena and described the behavior of such viruses. It is worth noting here that even though viruses
“are not part of the core system that arise by the interaction of the principles of UG and input during development, there is no reason to believe that they are not language and that the structures cannot interact with the language module”. In the words of Lasnik and Sobin, “[v]iruses comprise a subtheory of the Minimalist Program, distinct from the core system, though interactive with it”. Keeping L2A in mind, Lasnik and Sobin (p. 372, as cited in Bley-Vroman, 2009, p. 189) succinctly state:

Grammatical viruses mimic the original phenomenon, but
without complete success, since grammatical viruses are not of
the same descriptive power as the devices of the grammar
proper. Viruses lack generality and the possibility of normal
interactivity with the core grammatical system.

So, if the language system includes elements that are recruited outside the system, and if the language system itself is not working properly, this does not mean that these data processes (i.e. elements) are not available to the system.

In addition, increasing evidence from processing research studies suggests that for adult L2A “the syntactic representations…are shallower and less detailed than those of native speakers”. This, however, does not mean that native speakers do not employ shallow processing. Rather, it means that native speakers compute shallow representations, in addition to the deep grammatically defined representations. This account seems to resonate with Ulman’s (2001, p. 110, see above; Bley-Vroman, 2009, p. 192) suggestion that in L2, there is a “shift from procedural to declarative memory” of grammatical processing. Patches
can therefore be conceptualized as “words” or “rules” (but not the grammatically defined rules computed by the core grammatical system) that are processed in the declarative memory system (p. 192). To summarize the difference between L1A and adult L2A systems then, the following is a rough descriptive (or informal) sketch:

<table>
<thead>
<tr>
<th>Native language development</th>
<th>Foreign language learning</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Patch are periphery, used when the core system fails to work</td>
<td>Patches are central since the core system does not work properly</td>
</tr>
<tr>
<td>2 Data-processing mechanisms are recruited by the language acquisition system, but are not specific to language</td>
<td>Data-processing mechanisms recruited by the language acquisition system are used extensively outside the system since the system is not working properly, hence the mechanisms do not work well for language</td>
</tr>
<tr>
<td>3 The processing mechanisms of grammatical rules are computed mainly by deep processors, but sometimes by shallow processors</td>
<td>The processing mechanisms of grammatical rules are computed mainly (or perhaps exclusively) by shallow processors</td>
</tr>
</tbody>
</table>

(Bley-Vroman, 2009, pp. 192-193).

Cognitive Patch Theory is one of the most recent hypotheses in L2A, and to my knowledge it has not yet been empirically tested. The present study is therefore probably the first empirical study to put the hypothesis to the test. It is hoped that the results gained from this study will provide empirical evidence either for or against the posited hypothesis and in so doing advance L2 research. Any results that come out of the study could also have important implications for L2 pedagogy since such findings will provide empirical evidence either for or against the benefit of L2 instruction (e.g. explicit or implicit negative evidence vs. positive evidence only. The next chapter will set the scene for testing the Cognitive Patch Hypothesis by describing the participants, the choice of morphosyntactic test items including patches (or viruses), and the procedures followed in conducting the test.
Chapter 3:
The Method

3.1 Participants

The following is a table, which includes all the information about the participants of the study. The choice of the participants was guided by the fact that the morphosyntactic issues included in the grammaticality judgment test were often very subtle and only advanced ESL learners could be expected to deal with such fine-tuned judgments. Table 1: Biographical information about the participants of the study.

Table 1. A biographical information about the participants

<table>
<thead>
<tr>
<th></th>
<th>Advanced ESL grp</th>
<th>Native control gp</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
<td>20</td>
<td>16</td>
</tr>
<tr>
<td>L1 background</td>
<td>Different</td>
<td></td>
</tr>
<tr>
<td>Number of participants with knowledge of languages other than English and their L1s</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>Number of participants taking any of the standardized tests</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>Age range</td>
<td>Mostly 19-28</td>
<td>All 19-28</td>
</tr>
<tr>
<td>Length of residence in Canada</td>
<td>2 months-2.3 years</td>
<td></td>
</tr>
<tr>
<td>Number of participants reported receiving previous formal instruction in grammar</td>
<td>All but one</td>
<td></td>
</tr>
<tr>
<td>Educational background</td>
<td>University graduates and undergraduates</td>
<td>All grade 12 high school students graduating at the time of data collection</td>
</tr>
<tr>
<td>Gender</td>
<td>13 female; 7 male</td>
<td>9 female; 7 male</td>
</tr>
<tr>
<td>Level of Proficiency in ESL</td>
<td>The highest; Level 60</td>
<td></td>
</tr>
<tr>
<td>Educational institution where</td>
<td>A University in Toronto,</td>
<td>One of the high schools</td>
</tr>
</tbody>
</table>
3.2 Materials

1. A questionnaire was administered at the beginning of the test session. The items on the questionnaire included questions about the biographical information of the participants especially in reference to their personal experience with learning ESL. The questionnaire was included to make sure that the participants were all of the same level of proficiency in English and that all of them were of similar educational backgrounds. The questionnaire was also an attempt to make sure that all the participants had received previous formal instruction in grammar, a factor that is crucial to the present study since the study targets very subtle morphosyntactic features (for details about the questionnaire, see Appendix B, p.93).

2. The test included the following three tasks:

   (a) Task one was a timed grammaticality judgment task (TGJT). The term ‘grammaticality judgment task’ here does not mean that participants were asked to judge whether the test sentences were grammatically correct or not. Rather, they were asked to rate the sentences on a 7-point scale based on their intuitions or how they felt about them. However, the term will be used throughout, as it is the standard term used to describe this kind of tasks in the L2A research (see for example Ellis, 2005). In this task, a set of 60 randomized English sentences, 20 of them were possible, 20 impossible, and 20 patches were
administered. Table 1 in Appendix A (p. 73) specifies the test sentences together with their structural category: Task One (a timed grammaticality judgment task) was an attempt to tap into the participants’ implicit knowledge of the English language system. To this end, the test sentences that were selected for this study address some of the very subtle morphosyntactic categories of English, the knowledge of which would be highly unlikely to be derived from an explicit knowledge of the rules of English morphosyntax. The sentences created for the test address a number of morphosyntactic categories most of which are not the object of formal instruction in the ESL classroom. Rather, they reflect native speakers’ intuitive and unconscious knowledge of their language. Therefore, these morphosyntactic categories make an excellent testing ground for the reliability of the grammars acquired by the advanced ESL learners. Also included in the test are structural categories, which are assumed in the literature not to be part of the normal syntactic processes employed by native speakers. Rather, they are assumed to be linguistic patches. Therefore, including such patches in the test and analyzing the way advanced ESL learners respond to them should also answer the question whether or not advanced ESL learners demonstrate a performance similar to that of native speakers when dealing with such structures.

(b) Task two was a correction task, where the participants had each to correct all the sentences they rated 'impossible'. This meant that different participants would have different sentences to correct. This task was meant to test the participants metalinguistic knowledge of the English grammar system.
(c) Task three was a preference task (a multiple-choice task), where the participants were presented with 30 sets of sentences with four possible options in each set. The participants had to choose the best sentence in each set. This task was further meant to test the participants' metalinguistic knowledge of the morphosyntactic categories targeted in the test.

3. The participants were instructed to rate 60 English sentences in terms of their acceptability or naturalness on a seven-point scale, which ranged from (-3), representing ‘completely impossible’, to (+3), representing meaning ‘‘‘possible’. (For recommendations about the number of test sentences, see Gass & Mackey, 2005, pp. 50-51; for details about the scale used and the instructions given, see Bley-Vroman et al, 1988; Juff, 1996.). Fourty contrastive test sentences represent obervance versus violation of the normal syntactic categories targeted in the present study. In addition, twenty contrastive test sentences represent the linguistic patches targeted in the present study. The numbers of sentences in each category appear in the test had the arrangement in Table 1 below:

Table 2: Breakdown of test sentences

<table>
<thead>
<tr>
<th>Grammaticality</th>
<th>Possible</th>
<th>Impossible</th>
<th>patches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number</td>
<td>20</td>
<td>20</td>
<td>20</td>
</tr>
</tbody>
</table>

4. All the test sentences were created in the form of contrastive pairs. This means that half of the participants would receive one order and the other half would receive a different order. This was done in order to control for any effect of order on the performance of the participants. The sentences were randomized so that there was roughly a distance of 9
sentences between each form of a pair and its counterpart. This was not possible all the time; therefore two remaining sentences were placed farther apart from each other. An attempt was made to control for the length and lexical and syntactic complexity of each pair to preclude the possibility that the participants’ judgments were based on their intuitions about length and lexical and syntactic complexity. During the piloting stage, in some cases adding more words to some forms gave rise to new semantic and pragmatic nuance that seemed to influence the participants’ choices. Therefore, every possible effort was made to make sure that the forms of each contrasting pair had more or less the same length. However, in some cases, minor exceptions had to be made in order to provide learners with sentences that differed in syntactic structure. This was not possible all the time, since adding more words to some forms created new semantic and pragmatic factors that were found to influence the participants’ choice in the piloting stage. Sometimes, deleting words was not an option either, as this would have excluded the only syntactic difference, that made one form differ from its counterpart. Controlling for lexical complexity proved to be much easier than controlling for syntactic complexity. This was done by avoiding the use of some unfamiliar words, or words that were thought to be semantically problematic for the advanced ESL group. Controlling for syntactic complexity was minimal, and when done, every effort was made to ensure that each form was replaced with another that had the same syntactic properties (For a detailed account of the complexities involved in controlling for the length, lexical and syntactic complexity of items on a grammaticality judgment task, see Bley-Vroman, & Masterson, 1989.)
3.2.1 Grammatical Categories

Below is a description of the syntactic categories targeted together with some illustrative examples:

11. **Subjacency**: It is one of the UG principles that constrains the movement of items in a sentence so that items are not allowed to move just anywhere in the sentence. In English, items are only allowed to cross one bounding node at a time, where the bounding nodes are either a S(sentence) or a N(oun)P(hrase). Thus, sentences (1) and (2) below are ungrammatical, whereas sentence (3) is grammatical, as can be seen from their underlying representations.

(1)* What did Mary wonder whether John bought?

* What, [s did Mary wonder [CP whether [s John bought a]]]?

(2)* What did Mary believe the claim that John saw?

* What, [s did Mary believe [NP the claim [CP that [s John saw a]]]]?

(3) What did Mary think that John bought?

What, [s did Mary think [CP that [s John bought a]]]?

As can be seen from the underlying representations, sentence (1) is ungrammatical because the wh-word ‘what’ has crossed two bounding nodes (S, S) in violation of the subjacency principle. Sentence (2) is also ungrammatical because the wh-word ‘what’ has crossed three bounding nodes (S, NP, S) also in violation of the subjacency principle. Sentence (3), on the other hand, is grammatical because the wh-word ‘what’ has successfully
moved across one bounding node (S) into the intermediate position [ C(omplementizer) P(hrase)]\(^1\), Spec(ifier)], and then moving from that position into the highest landing site crossing one bounding node (S) at a time in accordance with the subjacency principle (White, 1989, pp. 23-24).

(1a) *Wh-islands*: Wh-words cannot be extracted from an embedded clause to the matrix clause (i.e. the main clause), as this would violate the Subjacency principle. This is because wh-words, when moving from their base position (i.e. original position or thematic position) into the highest position in the sentence, i.e. [Spec, CP], have to land in an intermediate [Spec, CP] position before they can reach their final hosting site. If, therefore, these intermediate [Spec, CP] positions are already taken by other wh-words, the transition from the base position to the highest [Spec, CP] position would be blocked, and this would render the sentence ungrammatical, as can be seen in the example below.

(4)* \([CP \text{What, } S \text{ does Mary want to know } [CP \text{ whether } [S \text{ John has already sold } s]]] \)\

Sentence (4) is ungrammatical because the wh-word ‘what’ crosses two bounding nodes (here two sentences) before landing in its final hosting position, thus violating the Subjacency Principle. The movement from the base position in the embedded clause could not have been made transitionally, since the intermediate position [Spec, CP] is already filled by another wh-word, ‘whether’ (Bley-Vroman, Felix, & Ioup, 1988, p. 10).

\(^1\)A complementizer, abbreviated as C is “a word such as that, for, whether which introduces a clause…and the whole phrase (complementizer plus clause) can be termed CP, a Complementizer Phrase” (Tallerman, 2005, p. 103).
(1b) *Complex NP islands*: These refer to the impossibility of extracting a wh-word from an embedded clause which is the complement of a nounphrase in the matrix clause, as this violates the Subjacency Principle. Consider the following example.

\[
(5)* [_{\text{CP}} \text{What, } [_{\text{S}} \text{ did Bill believe } [_{\text{NP}} \text{ the claim } [_{\text{S'}} \text{ that Carol had bought } _{\text{u}}]]]]?
\]

Sentence (5) is ungrammatical, since the wh-word crosses three bounding nodes at the same time, here taken to be two sentence (S’, S) and a nounphrase (NP), thus violating the Subjacency Principle (Bley-Vroman et al., 1988, p. 10).

(1c) *CP adjunct islands*: These refer to the impossibility of extracting a wh-word out of an adjunct CP, as can be seen below.

\[
(6)* \text{What } [_{\text{S}} \text{ did they leave the pub } [_{\text{CP}} \text{ before t finished with a tie}]]?
\]

In this sentence, the sentence is ungrammatical since it violates the Subjacency Principle, which prohibits the extraction of a wh-word out of two bounding nodes, here taken to be the two sentences, adjunct (CP) and (S) respectively.

12. *Case theory*: Case theory is a UG principle, which works at S-structure\(^2\), and posits the requirement that all lexical NPs (barring empty categories) receive their abstract case. On this account, verbs and prepositions assign cases to their objects, whereas subjects receive their case from INFL, the head of an inflectional phrase (IP), if the latter is tensed (i.e. a finite clause)\(^3\). This accounts for the differences in grammaticality between (7a) and (7a), and between (8a) and (8b) below respectively.

---

\(^2\)In UG accounts, S-structure is the derived structure, whereas D-structure is the base structure, the one where none of the items has yet undergone any movement (White, 1989, chapter 1).

\(^3\) In recent UG accounts, IP is another level in the underlying representation of a sentence, which replaced the level S in earlier UG accounts (White, 2003).
(7a) * It is likely John to be there.

(7b) John is likely to be there.

(8a) * Mary likes very much apples.

(8b) Mary likes apples very much.

Sentence (7a) is ungrammatical because the NP ‘John’ cannot receive its case from either the adjective ‘likely’ or the infinitival ‘to’, as neither of them is a case assigner. (7b), on the other hand, is grammatical because the NP ‘John’ moves from its thematic position to the beginning of the sentence, where it receives its case from INFL, a case assigner. In a similar vein, (8a) is ungrammatical since it violates the Ajacency Condition on Case Assignment (Chomsky, 1981a, 1986; Stowell, 1981, as cited in White, 1989, p. 24), which is a requirement that the NP receiving case must be adjacent to its case assigner. In other words, the NP ‘the apples’ in (8a) should be adjacent to its case assigner, the verb ‘like’ in order for the sentence to be grammatical (White, 1989, p. 24) (An alternative interpretation of the ungrammaticality of (8a) would be that the NP ‘apples’ is an argument, whereas ‘very much’ is an adjunct, and heads, i.e. verbs here, like to have their arguments close to them, see for example, Tallerman, 2005, pp. 98-100).

(2a) NP movement: This refers to the requirement that NPs, like wh-words, move from their base positions to their final landing host cyclically, i.e. stepwise. Consider the following example:

(9) [IP Mary, seems [IP t, to like Louise]].

---

4 The thematic position is the position of the item in the D-structure, where it receives its thematic role from the verb (White, 1989, 2003).
This sentence is grammatical because the NP ‘Mary’ moved from its base position in the lower [Spec, IP] to its final destination in the highest [Spec, IP]. This movement is required by Case Theory, where every NP has to have an abstract case assigned to it. Therefore, the NP ‘Mary’ has to move since it cannot get its case assigned in the base position because the infinitival ‘to’ is not a case assigner (Haegeman, 1994, pp. 306-309).

3. The AP\textsuperscript{5} -internal subject hypothesis: This hypothesis proposes that the [Spec, AP] position is the base (or thematic position) of the subject in sentences with copular ‘be’, indefinite subjects and existential ‘there’. The subject, according to the hypothesis, would thus move from its base position to the intermediate position [Spec, VP], since the higher position, [Spec, IP] is already taken by the existential ‘there’. Consider the following examples with their underlying representations:

(10a) There were three workers very tired.

\[
[\text{IP} \text{There} \text{[INFL were [VP three workers [V t [AP very [AP t tired]]]]]]}].
\]

(10b) *There were very three workers tired.

\[
* [\text{IP} \text{There} \text{[INFL were [VP very three workers [V t [AP t tired]]]]}]].
\]

Sentence (10a) is grammatical because the subject of the AP moves from its base position in the AP, [Spec AP], which is an A-position\textsuperscript{6} to the specifier of the VP, [Spec, VP], ‘three workers’, which is also an A-position, hence its grammaticality. In contrast, sentence

\textsuperscript{5}The abbreviation AP stands for the maximal projection Adjective Phrase, one of the lexical projections in the derivation of a sentence.

\textsuperscript{6}A-position is an argument position, i.e. a position, which can only be filled by an argument such as NPs and APs, which have a thematic role assigned to them by the verb. This contrasts with an A’-position, which is the position filled by elements such as the complementizers or wh-words, elements, which carry no thematic roles assigned to them by the verb (Haegeman & Gueron, 1999, pp. 220-227).
(10b) is ungrammatical because the subject in the AP does not move to [Spec, VP] since it is already taken by the adverb ‘very’, which is not in its proper position, rendering the sentence ungrammatical (Haegeman & Gueron, 1999, pp. 290-294).

4. The Binding Principle: This is a UG principle that works at all levels (S-structure, D-structure, L(ogical) F(orm)). It “constrains the relationship between various kinds of nounphrases”. Principle A of Binding theory is concerned with anaphors (reflexives and reciprocal pronouns) and states, “an anaphor must be bound in its governing category (the smallest NP or S containing the anaphor and a governor of the anaphor). That is, an anaphor must be coreferential with a nounphrase (i.e. antecedent), which is contained in the same sentence as that where the anaphor itself is contained. This accounts for the difference in grammaticality between (11a) and (11b), where the governing categories are bracketed for ease of illustration.

(11a) * John, said [s that Fred hurt himself,]

(11b) John said that [s, Fred, hurt himself,]

As can be seen above, (11a) is ungrammatical, as the anaphor ‘himself’ is governed by the NP ‘John’, outside of its governing category in violation of Principle A of Binding theory. In contrast, sentence (11b) is grammatical because the anaphor is governed by the NP ‘Fred’ which lies within its governing category, namely the sentence in which the anaphor is contained (White, 1989, pp. 26-27).

5. The Empty Category Principle: The Empty Category Principle (ECP) is a UG principle, which states that “a trace must be properly governed”, where proper government
entails that “\( \alpha \) properly governs \( \beta \) if \( \alpha \) governs \( \beta \) and \( \alpha \) is a lexical category (N[oun], V[erb], A[djective], P[reposition])”. Two syntactic phenomena can be explained through the ECP, namely the subject/object asymmetry and the behavior of the complementizer ‘that’ in English.

(5a) Subject/object asymmetry and the that-trace effect: In English, the complementizer ‘that’ is optional when the extracted wh-word is in an object position in the embedded sentence; whereas it is obligatory when the extracted wh-word is the subject of the embedded sentence. This is often referred to as the that-trace effect. Consider the following sentences in (12).

(12a) Who, do you think that Mary met t, yesterday?
(12b) Who, do you think Mary met t, yesterday?
(12c) Who, do you think t, arrived yesterday?
(12d) * Who, do you think that t, arrived yesterday?

As can be seen above, sentences (12a) and (12b) are both grammatical and the complementizer ‘that’ is optional. This results from the fact in both cases, the trace (t) of the wh-word is properly governed by the lexical governor, namely the verb ‘met’, a lexical category, in accordance with the ECP stated above. This is not the case in (12c) and (12d), where the wh-word was extracted from the subject position of the embedded clause, and where only the former is grammatical. Thus, while the trace in (12c) is properly governed by the antecedent wh-word, ‘who’, which is in the appropriate structural position in relation to the trace and which is coindexed with it; the same is not true for the sentence (12d). This is due to the fact that the trace in this sentence is not properly governed, as the antecedent wh-
word, ‘who’ is not in the appropriate structural position in relation to the trace (White, 1989, pp. 27-28)

(5b) **Superiority:** Consider the following two sentences:

(13) I can’t remember who t did what.

(14) * She forgot what who said t.

In (13), ‘who’ moves syntactically to the specifier position (Spec) of the complementizer phrase (CP), and the sentence is grammatical since the trace is properly governed lexically by the verb ‘did’. In (14), on the other hand, ‘what’ moves syntactically to [Spec, CP]; and the empty category is properly governed by the verb ‘said’. However, the sentence is ungrammatical, since the trace is not properly governed by the antecedent ‘what’ because the latter is not in the appropriate structural position in relation to the trace (Bley-Vroman et al., 1988, p. 11).

6. **The Head-Movement Constraint (HMC):** This is a constraint, which forbids a head to bypass another head position. Consider the following examples together with their underlying representations:

(15a) You could have done this.

(15b) Could you have done this?

\[ \text{[AGR}_A \text{GR}_A \text{ould}_L \left[ \text{TP}_T \text{t}_T \left[ \text{VP}_V \left[ \text{have}_I \text{done}_P \text{this} \right] \right] \right] \]?

(15c) * Have you could done this?

* \[ \text{[AGR}_A \text{GR}_A \text{have}_B \left[ \text{TP}_T \text{ould}_L \left[ \text{VP}_V \left[ \text{have}_I \text{t}_b \text{done}_P \text{this} \right] \right] \]?


As can be seen from the underlying representations above, sentence (15b) is grammatical since the highest auxiliary ‘could’, which is a head occupying the head position [TP, T’], moves to another head position, [AGRP, AGR]. In contrast, sentence (15c) is ungrammatical since the auxiliary ‘have’, which occupies the head position [VP, V] moves to the highest head position [AGRP, AGR] bypassing a potential head position, [TP, T’], which is already occupied by the higher auxiliary ‘could’, thus violating the (HCM) (Haegeman & Gueron, 1999, p. 332).

7. **Whether/if asymmetry**: In terms of their position in the underlying representation of the sentence, ‘if’ occupies a head position in C⁰, whereas ‘whether’ is a maximal projection occupying the position [Spec, CP]. This accounts for their selectional properties. Thus, while ‘if’ selects only finite clauses as its complement, ‘whether’, being a maximal projection, lacks the capacity to select its complements, and can therefore be combined with both finite and non-finite clauses as its complement. Consider the following examples:

(16a) I wonder if/whether I should go.

(16b) I wonder * if/ whether to go.

As can be seen from (16a), ‘if’ selects only a finite clause as its complement ‘I should go’, a capacity that stems from its position as a head, whereas this is not available for ‘whether’, which can be used with both finite and non-finite clauses as its complement ‘I should go/ to go’, since it is not a head, hence, unable to select its complement (Haegeman & Gueron, 1999, p. 319).
8. Finite/non-finite asymmetry: In English, auxiliaries move past the negation marker ‘not’ obligatorily in finite clauses, optionally in non-finite clauses. In non-finite clauses, lexical verbs never appear to the left of ‘not’.

Consider the following examples:

(17a) Thelma is not invited.
(17b) *Thelma not is invited.
(19a) Thelma does not talk.
(18b) *Thelma talks not.
(19a) Not to be invited is sad.
(19b) ?To be not invited is sad.
(20a) Not to get caught in the rain was a pure luck.
(20b) *To get not caught in the rain was a pure luck.

Sentences (17a) and (17b) are finite clauses. Therefore, the movement of the auxiliary ‘is’ across the negator ‘not’ is obligatory, as can be seen from the grammaticalness of the former but not the latter. Sentences (18a) and (18b) are also finite clauses. Therefore, the lexical verb ‘talk’ cannot move across the negator ‘not’ as can be seen from the grammaticalness of the former but not the latter. On the other hand, sentences (19a) and (19b) are non-finite clauses, where the movement of the auxiliary ‘be’ is optional, as can be seen from the somewhat acceptable status given to (19b) by speakers of some varieties of English. This is not the case with lexical verbs in non-finite clauses, where lexical verbs can
never cross the negator ‘not’ on their way to I(inflection) in IP, as can be seen from ungrammaticalness of (20b) (Haegeman & Gueron, 1999, pp.306-308).

9. Negative constituents vs. focus items asymmetry:

In English, negative constituents have the capacity to trigger subject-verb inversion; focus items do not. Consider these examples:

(21a) I promise that on no account will I write a paper during the holidays.
(21b)* I promise that on no account I will write a paper during the holidays.

Sentence (21a) is grammatical since the presence of the negative constituent ‘on no account’ has triggered the subject-verb inversion in the embedded clause ‘I will write a paper during the holidays’. On the other hand, sentence (21b) is ungrammatical since there is no subject-verb inversion triggered by the presence of the negative constituent (Haegeman & Gueron, 1999, p. 339-341).

In English, negative constituents also trigger the use of negative polarity items (e.g. anyone, anything), as can be seen from the difference in grammaticality in the following examples:

(22a) I promise that under no circumstances will I do anything wrong.
(22b)* I promise that under no circumstances will I do something wrong.

Sentence (22a) is grammatical because the preposed negative constituent ‘under no circumstances’ turns the force of the clause into a negative one, hence, triggers the use of the negative polarity item ‘anything’. On the other hand, sentence (22b) is ungrammatical since
the lexical item ‘something’ is not a negative polarity item (Haegeman & Gueron, 1999, p. 341).

The Head Movement Constraint (HMC) also predicts that the sequence ‘topic+ \textit{wh} constituent’ is ungrammatical, whereas the sequence ‘relativization + topicalization + negative constituent’ is grammatical, and the prediction is borne out, as can be seen in the different grammaticality of the examples below:

(23a) *This is a movie during the holidays which I would never watch.

"[CP[\textit{C}_C [TOPP during the holidays[?which[FOCP I would never watch]]]]]."

(23b) This is a movie which during the holidays never would I watch.

"[CP which [C [TOPP during the holiday[\textit{TOP}[+TOP[FOCP never[FOC [FOC would…]]]]]]]]."

Sentence (23a) is ungrammatical since the embedded relative \textit{wh}-word ‘which’ is not in its highest CP position, [Spec, CP], its proper position in the underlying representation of the sentence, and the sequence, topic ‘during the holiday’ + \textit{wh}-word’ is ungrammatical. In contrast, sentence (23b) is grammatical since the embedded relative \textit{wh}-word ‘which’ is in its proper position, [Spec, CP], and the sequence relativization + topicalization + negative constituent is grammatical (Haegeman & Gueron, 1999, pp. 346-347).

10. \textit{seem/copula be} asymmetry: There is another asymmetry in syntax between ‘seem’ and copula ‘be’. Thus, while the verb ‘seem’ is a lexical verb that can assign a theta role to a prepositional phrase (PP) complement; copula ‘be’ cannot. This can be seen from the difference in the grammaticality of the following two examples.

(24a) Jeff seems to me quite happy with his new job.

\footnote{The projection Focus Phrase [FOCP] is another maximal projection.}
(24b) * Jeff is to me quite happy with his new job

3.2.2 Extragrammatical Categories: Linguistic Patches

Linguistic patching refers to those cases where the grammar learning mechanism fails for some reason. The learner, not being able to form a generalization or a pattern that can cover all the relevant cases, resorts to forming a patch, that is, subsidiary principles other than the language faculty in order to handle such cases. Different speakers form different patches. Consider the following examples, where native speakers of English vary as to what the correct form is:

(1) ??Are/is (either) John or his parents here?
(2) ??Are/is (either) John’s parents or his wife here?

As can be seen in these examples, native speakers of English do not agree as to the correct form of the copula ‘be’ to be used. In addition, individual native speakers of English are not certain in their behavior regarding situations as the above, that is, they show inconsistency of behavior (Bley-Vroman, 2009, p. 187-190).

Linguistic patches have the following inherent characteristics:

(1) Lexical specificity: A patch strongly prefers particular lexical items, something that does not apply to normal syntactic processes, which are “normally lexicon-neutral and category-neutral.” (Sobin, 1997, p. 329). Consider the following examples:
(3a) My girlfriend and I went to see a movie last night.
(3b) ? My girlfriend and me went to see a movie last night.
Thus, while native speakers of English would select the nominative case in the structure (NP and I/me), they would do the opposite in a similar structure, exemplified in the following pair of sentences:

(3c) My girlfriend and him went to see a movie last night.

(3d) ? My girlfriend and he went to see a movie last night.

As can be seen from the examples above, native speakers of English normally select the accusative rather than the nominative case in the structure (NP and he/him) even though in both (3c) and (3d) the pronoun conjoins an NP in a subject position.

Now consider the following pair of sentences:

(4a) She and her friend went skiing.

(4b) ? I and my friend went skiing.

The same pattern is also revealed here. Thus, while the majority of native speakers of English select the nominative case in the conjunct (she and NP), they would do the opposite in the structure (I and NP) even though both structures are very similar in that the pronoun in both is being combined with an NP in subject position. Thus, it appears that the patch, which mimics the normal syntactic processes, selects the nominative right conjunct after ‘and’ as always the first person singular pronoun subject ‘I,’ but, which also seems to be generalized to ‘and PRNOUN’, hence the less consistent choice of ‘she’ in (She/her and NP). Thus, patch theory makes use of two rules here, the “…and I” rule, and the “that she…rule”, where the first rule selects the nominative conjunct on the right side of ‘and’, whereas the second rule
selects the nominative conjunct on the left side of ‘and’ (Quattlebaum, 1994, as cited in Sobin, 1997, p. 329).

(2) Directionality and adjacency: The “...and I” rule exhibits directionality since it assigns the nominative case to the right conjunct, but not to the left conjunct.

Consider the following two pairs of examples:

(5a) They invited my friend and I to the party.
(5b) ?They invited my friend and me to the party.

(6a) During the meeting, they listened to my colleague and I.
(6b) ?During the meeting, they listened to my colleague and me.

The majority of native speakers of English would select the nominative case after ‘and’ in the positions of object (5) and object of preposition (6). However, preceding ‘and’, only accusative first and third person pronouns are strongly selected (e.g., ...Verb she/\textit{her} and NP; Verb I/\textit{me} and NP; Prep she/\textit{her} and NP; Prep I/\textit{me} and NP, as can be seen in the following pairs of sentences:

(7a) They invited me and my friend to a party.
(7b) ???They invited I and my friend to a party.

(8a) I enjoyed listening to her and her friends.
(8b) ??? I enjoyed listening to she and her friends.

(9a) The manager spoke highly of me and my colleagues at the meeting.
(9b) ??? The manager spoke highly of I and my colleagues at the meeting.
These examples demonstrate that there is directionality in the use of the “...and I” rule (Quanttulebaum, 1994, as cited in Sobin, 1997, p. 329). The effect of adjacency can be seen by the way native speakers of English rate the following two structures:

(a) She/her and NP.
(b) NP and he/him.

Native speakers of English frequently select the nominative case in the first structure (a) and much less frequently select the nominative case in the second structure (b) (Quanttulebaum, 1994, as cited in Sobin, 1997, p. 329). Thus, the stronger tendency towards choosing the nominative case of the pronoun in the first structure can be accounted for by the fact that the pronoun is adjacent an underlying complementizer ‘that’ in the first but not the second structure. The selection of the nominative in the second structure can be attributed to two sources: the first might be a generalization of the” “...and I” rule to include pronouns other than the first pronoun ‘I’. The second source might be to relax the adjacency requirement on the “that she...” rule. Hence, the nominative is selected less frequently in the second structure than in the first (Sobin, 1997, p. 330).

(3) Overextension: The construction ‘between you and I’ that is often selected by native speakers of English must come from a source other than the normal case assignment, since prepositions only allow pronouns in the accusative case according to Case theory. One way to account for such a selection would be to consider it as an overextension of the “...and I” rule (Sobin, 1997, p. 331).

(4) Underextension: A patch is unable to affect all the items that are affected by the
normal grammatical process it mimics. Thus, while the normal case assignment process (i.e. based on Case theory) affects all the NP conjuncts of a coordinate NP (as in German, for example), the patch mimicking such a mechanism is unable to do so (Sobin, 1997, p. 331).

Nonlocality: Insensitivity to nonlexical hierarchic constituents: Consider the following two examples:

(10a) For Mary to be the winner and [SC I the loser] is unfair.

(10b) For Mary to be the winner and [INFLP me/??I to be the loser] is unfair.

As can be seen from the examples above, the nominative ‘I’ in (10a) is possible as a generalization of the nonlocal “…and I” rule. This is so because the pronoun does not seem to be sensitive to the nonlexical (i.e. nonphonetic) higher constituent (i.e. SC= small clause). However, this is not the case in (10b), where the nominative pronoun ‘I’ does not seem to be acceptable due to the influence of the higher constituent (INFLP or TP), which is lexically (i.e. phonetically) projected by the item ‘to’ (Sobin, 1997, P. 331).

3.3 Procedures

The following steps were followed during the test administered to the advanced ESL and native-speaker groups:

1. The native participants were all grade 12 high school students graduating at the time of data collection. All the participants were studying at one of the high schools located in Toronto, Ontario, Canada. The researcher posted an advertisement on the social network ‘Facebook’ looking for participants who were born in Canada and whose dominant language at home is English. The researcher was then contacted by one of the participants who then
helped to arrange for her classmates to take part in the study in exchange for a sum of $20 dollars paid for each participant. Only potential participants who met the above requirements were then selected to take part in the study. A biographical questionnaire was administered only to the ESL group, which included questions about the participants’ history of learning English. The native-speaker participants were given a time limit of 6 minutes to finish the first task. This time limit was based on the average time it took native speakers to finish the task during the piloting stage of the study. The advanced ESL group had a time limit of 7.2 to finish the first task. This time limit was determined based an average time of 6 minutes, the average it took native speakers to finish the task during the piloting stage of this study, plus extra 20% of the time for each test sentence (for details about this decision for the timing of the task, see Ellis, 2005, p. 156). The participants were first asked to read the instructions. Once the participants finished reading the written instructions, the researcher explained the task orally to make sure that the participants knew exactly what was required of them in this particular task. The participants were given pens to do this task in an attempt to prevent them from changing their answers (cf. Gass & Mackey, 2005, p. 51).

2. Once all the participants finished the first task, they were asked to go back to the task again and correct all the sentences that they rated as unacceptable or impossible. They were orally instructed to rewrite the sentences so as to make them acceptable. The participants were asked to try their best to fix the form of the sentences while keeping the meaning as much as was possible. This task was important in order to see whether the participants knew exactly what was wrong with the ungrammatical sentences or whether their
correct judgments were simply the result of wrong interpretations of the sentences’ ungrammaticality (cf. Gass & Mackey, 2005, pp. 48-50). There was no time limit for this task, as it was meant to tap into the participants’ explicit knowledge of the rules of English grammar.

3. Once the second task was finished, the participants were given the third task (a preference task) in which they were asked to select only one out of four possible options in a set of 30 English sentences. They were asked to choose the option that represented their best choice, i.e. the one option that they liked more than all the others. The sentences were also given to the participants in two different orders to control for any effect order may have on the performance of the participants. The sentences were randomized in terms of grammaticality so that any guessing strategy to pick the right choice would fail. There was no time limit for this task, as it was meant to tap into their explicit knowledge of the English language system.

The procedure followed with the NSs control group was the same as above except for two details. The first is that with this group the time limit for the first task (i.e. grammaticality judgment task) was only 6 minutes rather than 7.2 minutes, as was the case with the ESL group. This decision was guided by the average time it took 4 native speakers of English to finish this task during the piloting stage. The second detail in the procedure used with this group is that the participants did not fill out a biological questionnaire, as this was not needed for the NS control group.
Following Bley-Vroman (2009), the following hypotheses were tested:

1. The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will be unreliable, where reliability is measured as the degree to which the acquired grammars match the L2 grammar.

2. The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will not converge to the L2 grammar, where convergence is measured as the relationship among the acquired grammars.

3. The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will be characteristic of patches, where patches are processed by extragrammatical principles independent of the normal syntactic processes.

This chapter has described the participants who took part in the study, the procedures followed in the development of the target test items including all the morphosyntactic categories and their exemplar test sentences and the characteristics of patches, as described by Sobin (1997). In the next chapter, the data analysis is described and the results are reported.
4.1 Testing The Reliability and The Convergence Hypotheses in Task 1 (TGJT)

In this section, the analysis of the general performance of both the advanced ESL group and the native control group on task 1 (the timed grammaticality judgment task) is described. Only the 20 ‘possible’ (observing the syntactic categories targeted) and the corresponding 20 'impossible' (violating the syntactic categories targeted) sentences will be analyzed here. The aim is to test the first hypothesis (i.e. the reliability hypothesis) and the second hypothesis (i.e. the convergence hypothesis).

HYPOTHESIS ONE: The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will be unreliable, where reliability is measured as the degree to which the acquired grammars match the L2 grammar.

HYPOTHESIS TWO: The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will not converge to the L2 grammar, where convergence is measured as the relationship among the acquired grammars.

For the purposes of analysis, a participant’s rating of a given item was considered as correct when possible and impossible items are judged accurately. Although participants were also given the choice ‘I don’t know’, this judgment counted as wrong on a par with a 'possible' rating for an impossible item and an ‘impossible’ rating for a possible item. This decision was based on the idea that such a choice reflects ‘uncertainty’ on the part of the participant, which, for the purpose of this study, was interpreted to mean ‘lack of knowledge’
(for similar decisions, see Bley-Vroman et al., 1988, p. 16). A t-test was run to compare the overall performance of the two groups. Table 1 shows the results of the statistical analysis (See Table 3 in Appendix A, p.86 for a detailed analysis of the overall performance of the participants of both groups).

Table 1. The overall performance of the ESL group and the NS group on Task 1 (TGJT).

<table>
<thead>
<tr>
<th>Task 1 (TGJT)</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESL group</td>
<td>20</td>
<td>63.75</td>
<td>9.07</td>
<td>2.02</td>
</tr>
<tr>
<td></td>
<td>NS group</td>
<td>16</td>
<td>75.5</td>
<td>7.79</td>
<td>1.94</td>
</tr>
</tbody>
</table>

As can be seen from the table above, the native speakers’ average score was 75.5%, and the ESL participants’ average score was 63.75% (with a relatively small SD score in both cases) (ESL participants: SD= 9.07; NS participants= 7.78). The difference between the scores is statistically significant: t (34) = -4.11, p < .001. However, the ESL participants’ score was significantly better than chance (assuming 50% to be chance). The 95% confidence interval for the difference between the means ranged from –17.56 to –5.94. Cohen’s d effect size statistic was 1.38 indicating a large effect.

The results gained from the first task indicate that the first hypothesis cannot be maintained as it stands since the performance of the advanced ESL participants was significantly better than chance (M = 63.75%); yet not as good as that of the NS participants (M = 75.5%) since there was a statistically significant difference between the general
performance of the two groups (p< . 001). Based on the results form this task, it can therefore be said that the hypothesis is partially disconfirmed.

With regard to the second related hypothesis, the results obtained from this task show a similar picture. Here also, the results show that acquired grammars of advanced ESL participants partly converge to the L2 grammar. This is because their performance is significantly better than chance (M = 63.75%); yet their convergence to the L2 grammar is not as good as that of the NS participants. The second hypothesis can therefore be partially disconfirmed.

4.2 Testing The Reliability and The Convergence Hypotheses in Task 2 (the correction task)

This task was designed to further test the same two hypotheses stated earlier. In this section, I will analyze the 20 impossible test sentences together with the corrections offered by the participants of both groups. The order of the syntactic categories and the sentences that violate them will follow the one used in the Method chapter (ch.3). As before, the aim is to test the reliability and the convergence hypotheses posited in the third chapter. A participant’s response to this task was considered correct only if they marked the ‘impossible’ sentence as such and provided a correction, which does not violate the syntactic category targeted; a participant’s response was considered incorrect otherwise. A t-test was run to compare the overall performance of the two groups. Table 3 shows the results of the statistical analysis.
Table 2. The overall performance of the ESL group and the NS group on task 2.

<table>
<thead>
<tr>
<th>Group</th>
<th># of test sentences</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task 2</td>
<td>The correction task</td>
<td>ESL</td>
<td>20</td>
<td>39</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mean</td>
<td>23.42</td>
<td>5.23</td>
</tr>
<tr>
<td></td>
<td></td>
<td>NS</td>
<td>20</td>
<td>58.5</td>
</tr>
</tbody>
</table>

As can be seen from the table above, the native speakers’ average score was 58.50%, and the ESL participants’ average score was 39%. The difference between the scores is statistically significant: $t(38) = -2.49, p < .05$. The 95% confidence interval difference of the means ranged from $-35.33$ to $-3.67$. The Cohen’s $d$ effect size was 1.27 indicating a medium effect. The ESL participants’ score was significantly better than chance (assuming 25% to be chance) (see Table 5 in Appendix A, p. 88 for a detailed analysis of the performance of both groups on task 2).

Based on the results gained from this task, the interlanguage grammars of the advanced ESL participants were reliable (39%, assuming a chance level of 25%). However, the interlanguage grammars of the ESL participants was not as good as that of the NS participants since there was a significant difference between the performance of the two groups ($p < .05$).

Based on the results gained from this task, the interlanguage grammars of the advanced ESL participants did converge to the L2 grammar since their performance was much better than chance (assuming a chance level of 25%). The second hypothesis is
therefore partially disconfirmed. See Table 4 in the Appendix for a detailed analysis of the performance of both groups on task 2 (the correction task).

4.3 Testing The Reliability and The Convergence Hypotheses in Task 3 (the preference task)

In this section, I will analyze the overall performance of both the advanced ESL group and the native comparison group on task 3 (the preference task or multiple-choice task). Only the 20 possible and the corresponding 20 impossible sentences will be analyzed here. Here also, the aim is to test the reliability and the convergence hypotheses posited above. Each participant was assigned a score, which is the percentage of the sentences chosen correctly. A participant’s preference was considered correct when the sentence chosen was the only ‘possible’ option in the four options given for each set. A participant’s preference was considered incorrect when the participant chose any sentence other than the only ‘possible’ sentence given in each set. A t-test was run to compare the overall performance of the two groups. Table 2 shows the results of the statistical analysis.

Table 3. The overall performance of the ESL group and the NS group on task 3.

<table>
<thead>
<tr>
<th>Task 3 The preference task</th>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>Std. Deviation</th>
<th>Std. Error Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ESL group</td>
<td>20</td>
<td>67.5</td>
<td>12.82</td>
<td>2.86</td>
</tr>
<tr>
<td></td>
<td>NS group</td>
<td>16</td>
<td>84.68</td>
<td>14.39</td>
<td>3.57</td>
</tr>
</tbody>
</table>

As can be seen from the table above, the native speakers’ average score was 84.68%, and the ESL participants’ average score was 67.50%. The difference between the scores is statistically significant: t (34) = -3.80, p < .05. The 95% confidence interval difference of the
means ranged from –26.39 to –7.83. Cohen’s d effect size was 1.27 indicating a large effect. The ESL participants’ score was significantly better than chance (assuming 50% to be chance) (see Table 4 in Appendix A, p. 87 for a detailed analysis of the performance of both groups on task 3).

The results gained from this task indicate that the first hypothesis cannot be maintained as it stands since the ESL participants performed significantly better than chance (M= 67.50); yet not as good as the NS participants did (M= 84.68) since the difference in the overall performance of both groups was statistically significant (p<.05). Therefore, it can be said that the first hypothesis is partially disconfirmed.

The results obtained from this task also indicate that the second hypothesis cannot be maintained as it stands since the acquired grammars of advanced ESL participants were converging to the L2 grammar. This is because their performance is better than chance (M= 67.50); yet their convergence to the L2 grammar is not as good as the native speakers’ convergence to the L2 grammar (84.68), since the difference in the overall performance of both groups was statistically significant (p<.05). Therefore, the second hypothesis is partially disconfirmed.

4.4 The ESL Group: Testing The Patching Hypothesis

This task was designed to test the third hypothesis:

HYPOTHESIS THREE: The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will be characteristic of patches, where patches
are processed by extragrammatical principles independent of the normal syntactic processes.

The performance of advanced ESL students in task 1 and 2 (i.e. the rating task and the correction task) in the way they treated the 20 linguistic patches was analyzed against the characteristics of patches (or viruses) as described by Sobin (1997). A participant’s performance was considered patchy if in their judgments of each linguistic patch they demonstrated a behavior which matched any of the characteristics of linguistic patches described in the Method chapter (Ch. 3). The overall performance of advanced ESL participants on linguistic patches is characteristic of patches (M= 83%).

The performance of advanced ESL participants on linguistic patches is very similar to the native speakers’ performance on linguistic patches, as has been documented in the literature (see Table 6 in Appendix A, p. 91 for a detailed analysis of the performance of the ESL participants on linguistic patches). Based on the results of this test, this hypothesis cannot be maintained as it stands since the results only show that the performance of advanced ESL participants on linguistic patches is characteristic of patches, and that their behavior is very similar to that of native speakers who demonstrate a similar behavior when it comes to patches. Therefore, the hypothesis is disconfirmed.
Chapter 5:
Discussion and Conclusion

5.1 Discussion

This study aimed to investigate the following three hypotheses:

(1) The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will be unreliable, where reliability is measured as the degree to which the acquired grammars match the L2 grammar.

(2) The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will not converge to the L2 grammar, where convergence is measured as the relationship among the acquired grammars.

(3) The interlanguage grammars of advanced L2 learners of English in the area of morphosyntax will be characteristic of patches, where patches are processed by extragrammatical principles independent of the normal syntactic processes.

The results gained from the data analysis partially disconfirm the first hypothesis since the acquired grammars of the advanced ESL participants were reliable in that the overall performance of the participants was better than chance. However, there was a statistically significant difference between the two experimental groups in favor of the NS group in terms of whether the acquired grammars match the target grammar or not on the three tasks: the TGJ task, the correction task and the preference task.

The results of this study also partially disconfirm the second hypothesis since the acquired grammars of the advanced ESL participants do converge to the L2 grammar. However, their
convergence to the L2 grammar is not as good as that of the native speakers since there was a statistically significant difference between the acquired grammars of the two groups (in favor of the NS group) on the three tasks of the experimental design. Thus, while the acquired grammars of advanced ESL learners converge in that they resemble one another, they do not converge to the L2 grammar in the same way the grammar of the native-speaker participants does.

The results of this study disconfirm the third hypothesis since the performance of the advanced ESL group has been shown to be not more patchy than that of native speakers. However, the acquired grammars behave in a way similar to the grammar of native speakers on linguistic patches in that both grammars display patchiness in areas where normal syntactic processes do not seem to work.

The results also reveal that even though there was a statistically significant difference in favor of the NS group on the second task (the correction task), both experimental groups did much better on the first task (TGJ) than on the second (the correction task). One interpretation of this finding is that the second task required that the participants use their metalinguistic knowledge of the syntactic categories being investigated. Taking into account the fact that most (if not all) of the targeted syntactic categories (barring linguistic patches) are exemplars of unconsciously acquired principles that are not accessible to the conscious knowledge of the NS group, the poorer performance of the NS group on this task compared to their much better performance on the first task, (which was aimed at indirectly tapping into their unconscious knowledge of the syntactic categories targeted) should come as no surprise.
On the other hand, the fact that even the advanced ESL group performed poorly on the second task compared to their performance on the first task might be explained by the fact that most of the syntactic categories targeted in this study are not topics for formal instruction in L2 English. In other words, any metalinguistic knowledge that the advanced ESL participants have about English would be of little to no help on the tasks of this study.

5.2 Conclusions

The results of this study partially disconfirm the cognitive patch hypothesis (Bley-Vroman, 2009) in that they show that the interlanguage grammars of the advanced ESL learners are reliable even though not as reliable as that of native speakers. Also, while it is true that the interlanguage grammars of advanced ESL learners converge to the L2 grammar, they do not seem to converge to the target grammar as successfully as the acquired grammar of the native speakers does. Yet, while the behavior of advanced ESL learners is characteristic of patches is confirmed in so far as linguistic patches are concerned, the hypothesis that the overall performance of advanced ESL participants in the area of morphosyntax is characteristic of patches is disconfirmed.

Building on the results of this study, it seems that the cognitive patch theory (Bley-Vroman, 2009) cannot be maintained, and that another hypothesis might capture the similarities and differences between L1A and adult L2A. To be sure, the results of the present study fall more neatly into the predictions made by other generative approaches to SLA, particularly White (2003) and Herschensohn (2000, 2009), where the claim is that the difference between the L1 and L2 acquisition processes is quantitative rather than
qualitative. The results of this study have shown a more complex picture than predicted. The performance of advanced ESL participants was better than random; the participants have demonstrated ability to make very subtle morphosyntactic judgments, which cannot be simply attributed to any amount of exposure to the L2 input. In other words, their interlanguage grammars, while diverging from the grammar of native speakers, can be said to fall within the bounds laid out by UG. At the same time, however, there was a statistically significant difference between their acquired grammars and that of the native speakers of English on all three tasks employed in this study. This leaves us with one interpretation, namely, that while convergence to the L2 grammar is possible, it is not guaranteed, a view that has been adopted by many UG researchers in SLA (see for example, Schwartz and Sprouse, 1994, 1996; Eubank, 1993/4, 1994; Vannika and Young-Scholten, 1994, 1996; White, 2003, as cited in White, 2003).

5.2.1 Pedagogical Implications

The results suggest that the learning mechanisms of the advanced L2 learners is reliable even though not as reliable as that of the native speakers. The results also suggest that the interlanguage grammars of advanced ESL participants converge to the L2 grammar even though not in the same way as the acquired grammar of native speakers does. The findings further suggest that the interlanguage grammar of advanced ESL participants is patchy, but only as far as linguistic patches are concerned. Therefore, the results as they stand, lend support to the view that L2 learners require both positive evidence in the form of natural input that is not manipulated and negative evidence in the form of recasts and clarification
questions as well as explicit error correction instruction. Also, the results indicate that more explicit and implicit input is required even at the highest level of L2 acquisition in the hope that such manipulated input triggers the kind of “awareness” required to boost the learning mechanisms of L2 learners and help them work more efficiently.

Thus, the fact that unlike child L1 learners, adult L2 learners are selective in the way they allocate their attentional resources indicates the importance of manipulating the input these learners receive since such manipulation in the form of FOF(s) would help to make certain properties of the L2 more “noticeable” to L2 learners.

It is also crucial to point out that linguistic theory especially in its latest incarnations (the Principles and Parameters/Minimalism) is most capable of accounting for language variation by attributing such variation to differences in the morphosyntactic properties of lexical items in different languages. It is the author’s belief, therefore, that pedagogy would fare much better should more emphasis be given to such morphosyntactic properties that make languages different.

5.3 Limitations of the Study and Suggestions for Future Research.

There are a number of limitations that can be pointed at about this study:

First, the experimental design of this study included using tokens of various syntactic categories. Yet, there were not equal numbers of tokens for each of the syntactic categories targeted. Thus, some categories had more tokens than others. A future study might therefore select only one of the syntactic categories investigated here with a large number of tokens so that any results gained from the study can therefore be generalizable.
Second, most of the patches included in the study were disjunct NPs in subject position. Another study might focus on the way native speakers and adult L2 learners respond to structures where the case of NPs is being manipulated. It would be interesting to see whether L2 learners, like native speakers, will exhibit a performance characteristic of patches in this area.

Third, the results of this study have shown that native speakers performed poorly on some syntactic structures (particularly CP adjunct islands, finite/non-finite asymmetry, and negative constituents as triggers of negative polarity items). Unfortunately, the number of tokens included in this study for each of these structures is not enough to make any generalizations. Therefore, it would be interesting to see how both groups compare in an experimental design where the focus is on these structures and where enough tokens are used so as to make generalizations possible.

Moreover, the linguistic patches included in this study include existential sentences with ‘there…’ in subject position. It could be illuminating to see how native speakers and advanced L2 learners behave in an experimental design where this kind of structure is contrasted with expletive constructions with ‘it’ as the subject of the sentence.

Finally, this study has looked at the performance of advanced ESL learners with different L1 backgrounds. It was therefore difficult to decide whether certain properties in the behavior of advanced ESL participants can be attributed to L1 transfer or to UG-constrained properties. Another study might look at the performance of advanced ESL learners who share the same L1 background. Such a study would make it possible to tease out properties of the
interlanguage grammar that can be attributed to L1 and those that are neither L1 nor L2 but still UG-constrained.
References


London: Routledge.


Appendix A
Table 1. The syntactic categories and test sentences.

<table>
<thead>
<tr>
<th>Test item</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Subjacency Principle</td>
<td></td>
</tr>
<tr>
<td>a. Wh-islands</td>
<td></td>
</tr>
<tr>
<td>1. Why [c.d] they wonder t whether she would leave the key in the car?</td>
<td>This sentence is grammatical because it observes the Subjacency Principle in that the wh-word moves from its base position in the matrix clause marked by the trace ‘t’ to the Spec, CP position of that same clause.</td>
</tr>
<tr>
<td>2. * [CP Where, [c.d] they wonder whether [IP2 she would leave the key t?]]</td>
<td>This sentence is ungrammatical because it violates the Subjacency Principle. The wh-word ‘where’ makes an illegitimate movement from its base position in the embedded clause to the highest position in the Spec, CP of the matrix clause crossing two bounding nodes, here taken to be two sentences [IP 1, IP 2].</td>
</tr>
<tr>
<td>b. Complex NP islands</td>
<td></td>
</tr>
<tr>
<td>3. [CP Why, [c.d] the fact [CP t, [C that [IP2 Morag put the ice cream on the table] [IP1 make you angry t?]]]]</td>
<td>This sentence is grammatical because it observes the Subjacency Principle. The wh-word ‘why’ moves from its base position marked by the trace ‘t’ to the highest Spec, CP position crossing one bounding node at a time. In other words, the movement takes</td>
</tr>
<tr>
<td>No.</td>
<td>Sentence</td>
</tr>
<tr>
<td>-----</td>
<td>----------</td>
</tr>
<tr>
<td>4. *</td>
<td>*CP Where, [IP1 did [NP the fact ...]]</td>
</tr>
<tr>
<td>5.</td>
<td>CP How, [IP did they leave the pub ...]]</td>
</tr>
<tr>
<td>6. *</td>
<td>*CPi How [IP did they leave the pub ...]]</td>
</tr>
</tbody>
</table>
moved from its base position to in the adjunct clause to the highest Spec, CP in the matrix clause crossing two bounding nodes, namely the CP adjunct phrase and the IP respectively. In other, the movement is not cyclical.

2. **Case Theory**

a. **NP Movement**

7. \[ [i_p \text{ It seems } [c \text{ that } [i_p \text{ John}, \text{} is likely } [i_p \text{} t, \text{} to \text{} go.]]]] \]

This sentence is grammatical because it observes Case Theory. The NP ‘John’ leaves its base position in the lower IP because the infinitival ‘to’ is not a case assigner. Therefore, the NP moves to ‘I’ in IP, to gets its nominative case assigned. The NP ‘John’ and its trace form a chain called A-chain, and case is assigned only to the head of the chain, namely the NP ‘John’.

8. * \[ [i_p \text{ John, seems } [c \text{ that } [i_p \text{} \text{} it is likely } [i_p \text{} t, \text{} to \text{} go.]]]] \]

This sentence is ungrammatical because it violates Case Theory. This is because the NP ‘John’ cannot get its case from the infinitival ‘to’, which is not a case assigner. Since the movement of the NP ‘John’ from the lowest IP to the highest IP has skipped another potential landing site, namely the subject position of the intermediate IP position, which is already, occupied by the expletive ‘it’. Thus, the
sentence is rendered ungrammatical. In other words, the movement of the NP ‘John’ has not been cyclical, thus also violating the locality condition according to which movement in general should be local (i.e. in steps).

9. [\text{IP} \text{It was thought} [c \text{ that [IP Mary would arrive late.]}] This sentence is grammatical because it observes Case Theory. The NP ‘Mary’ receives its nominative case in the lower IP position, where the head ‘I’ is a case assigner since the clause is finite.

10. * [\text{IP} \text{It was thought} [\text{IP Mary arriving late.}]] This sentence is ungrammatical because it violates Case Theory. The NP ‘Mary’ is in the lower IP, where the head ‘I’ is not a case assigner since the clause is non-finite.

11. \text{[IP John, seems [IP t_i to like Mary.]]]} This sentence is grammatical because it observes Case Theory. The NP ‘John’ in the lower IP leaves its position since the lower IP is non-finite, hence not a case assigner. It lands in the highest IP to get its nominative case assigned since this IP is finite, hence a case assigner.

12. * [\text{IP John, seems [c \text{ that [IP he, likes Mary.]]]}]] This sentence is ungrammatical because it violates both Case Theory and Theta theory. The fact that both the pronoun ‘he’ and its antecedent ‘John’ are
coreferential requires that both have one case and one theta role. Since the pronoun ‘he’ receives its nominative case and its theta role in the lower IP, the NP ‘John’ is thus left with no case and no theta role assigned to it. This is because case and theta roles are assigned to one and the same NP only once.

3. The AP-internal subject hypothesis

13. * [IP There [I were, [VP three students, [V t, [AP very [AP t, sick.]]]]]]]
   This sentence is grammatical because it observes the Adjective Phrase Internal Subject Hypothesis. The subject NP ‘three students’ moves from the AP position to the intermediate VP position since the highest IP position is taken by the expletive ‘there’.

14. * [IP There [I were, very three students [V t, [AP t sick.]]]
   This sentence is ungrammatical because it violates the AP-internal subject Hypothesis. The movement of the subject NP ‘three students’ to Spc, IP is blocked by the fact that the degree adverb ‘very’ does not modify NPs. Rather, it can only modify APs.

4. The Binding Principle

15. [IP Mary, believes [IP John, to expect pictures of himself,] to be on
   This sentence is grammatical because it observes the Binding Principle. The reflexive ‘himself’ is bound
sale next week. (i.e. c-commanded and has the same phi-features of
number, person, and gender) by an antecedent in its
own local clause. In other words, the sentence is
grammatical according to Condition A of the
Binding Principle.

16. [IP Mary, believes] [IP John, to expect pictures of herself,] to be on
sale next week.
The sentence is ungrammatical because it violates
Condition A of the Binding Principle. The reflexive
‘herself’ is not bound by an antecedent in its own
local clause, as can be seen from the indexes.

17. [IP John, considers] [IP Mary, to be too proud of herself,]]
This sentence is grammatical because it observes
Condition A of the Binding Principle. The reflexive
‘herself’ is bound by an antecedent in its own local
clause.

18. * [IP John, considers] [IP Mary, to be too proud of himself,]]
This sentence is ungrammatical because it violates
Condition A of the Binding Principle. The reflexive
‘herself’ is not bound by an antecedent in its own
local clause, as can be seen from the indexes.

5. The Empty Category Principle (ECP)
a. Superiority Principle

19. He wondered [CP who, [IP t, said what.]]
This sentence is grammatical because it observes the
Empty Category Principle (ECP). The wh-word
‘who’ moves from its base position in IP marked by
the trace ‘t’ to [Spec, CP], and the variable ‘t’ is
properly governed by its antecedent ‘who’ since the
20. *[CP [Spec, What, who, [c did [IP he wonder [ip t, said t_j,]]]]] This sentence is ungrammatical because it violates the ECP. The wh-word ‘who’ moves to [Spec, CP] and properly governs its variable ‘t,’ since it c-commands it. The wh-word ‘what,’ on the other hand, moves to adjoin ‘who’ in [Spec, CP’; yet the maximal projection CP has the properties of ‘who’, as shown by the fact that they are coindexed. Therefore, ‘what’ does not properly c-command its trace ‘t’.

b. Subject/object asymmetry

21. [CP Who, [c did [IP you wonder [CP whether [ip they had called [ip t_1, called t_2,]]]]]] This sentence is grammatical because it observes the ECP. The trace of the object ‘who’ in the lowest IP is properly governed by the lexical verb ‘call’, which assigns to its complement its theta-role. Therefore, the variable ‘t,’ while not properly governed by its antecedent ‘who’, it is properly governed by the lexical verb ‘call’.

22. * [CP Who, [c did [IP you wonder [CP whether [ip t, called t_1, called t_2,]]]]]] This sentence is ungrammatical because it violates the ECP. The variable ‘t,’ is not properly governed neither by its antecedent ‘who’ because the complementizer ‘whether’ in the lower [Spec, CP] position prevents ‘who’ from antecedent-governing.
its variable ‘ti’. Also, the variable ‘ti’ is not theta-
governed by the lexical verb ‘call’. In other words,
the variable ‘ti’ is not bound by neither the
antecedent nor the lexical verb.

c. That-trace effect

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>23. [\text{CP} \text{Who} [c \text{ do } [\text{IP} you think [\text{CP} ti [\text{IP} \text{ti will invite Mary?}]]]]]</td>
<td>This sentence is grammatical because it observes the ECP. The variable ‘ti’ in the lowest IP position is properly governed by another variable ‘ti’ in the lower [Spec, CP], and the latter is properly governed by its antecedent ‘who’.</td>
</tr>
<tr>
<td>24. * [\text{CP} \text{Who} [c \text{ do } [\text{IP} you think [ti that [\text{IP} \text{ti will invite Mary?}]]]]]</td>
<td>This sentence is ungrammatical because it violates the ECP. The variable ‘ti’ in the lowest IP position is not properly governed by the variable ‘ti’ in the lower [Spec, CP] since ‘that’ blocks such antecedent-government.</td>
</tr>
</tbody>
</table>

6. Head-Movement Constraint (HMC)

<table>
<thead>
<tr>
<th>Sentence</th>
<th>Analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>25. I promise [c that never [\text{IP} \text{will} \text{VP} I drink] on Sundays.</td>
<td>This sentence is grammatical because the auxiliary verb ‘will’ moves from ‘V’ to ‘I’, a valid movement.</td>
</tr>
<tr>
<td>26. * I promise that never [\text{IP} \text{will} \text{FocP on Sundays [\text{VP} I drink.]]}</td>
<td>This sentence is ungrammatical because it violates the Head Movement Constraint (HMC). The movement of the auxiliary verb ‘will’ from ‘VP’ to ‘IP’ is blocked by the existence of another head, namely ‘Foc’, which is the head of the focus phrase.</td>
</tr>
<tr>
<td>27. Why (<em>{C}) did (</em>{IP}) they (_t) invite Tom on Sunday?</td>
<td>‘on Sunday’. Therefore, the movement is blocked because it is not cyclical.</td>
</tr>
<tr>
<td><strong>This sentence is grammatical because the auxiliary ‘did’ moves from ‘I’ to ‘C’, a valid movement.</strong></td>
<td></td>
</tr>
<tr>
<td>28. * Why (<em>{C}) did (</em>{FocP}) they invite Tom on Sunday (_{IP})?</td>
<td>This sentence is ungrammatical because it violates the HMC. The auxiliary ‘did’ moves from ‘I’ to ‘C’ crossing another head, namely ‘Foc’ of FocP.</td>
</tr>
<tr>
<td><strong>7. Whether/if asymmetry</strong></td>
<td></td>
</tr>
<tr>
<td>29. I wonder whether or not they will come.</td>
<td>This sentence is grammatical because both ‘whether’ and ‘not’ are maximal projections; therefore, they can be coordinated.</td>
</tr>
<tr>
<td><strong>8. finite/non-finite asymmetry</strong></td>
<td></td>
</tr>
<tr>
<td>31. Not to get arrested for murder is a miracle.</td>
<td>This sentence is grammatical. This is because the lexical verb ‘get’ is to the right of the negative head ‘not’, in accordance with the grammatical rules of English.</td>
</tr>
<tr>
<td>32. * To get not arrested for murder is a miracle.</td>
<td>This sentence is ungrammatical because the lexical verb ‘get’ crossed the negative head ‘not’ in violation of the grammatical rules of English.</td>
</tr>
<tr>
<td><strong>9. Negative constituents/focus items asymmetry</strong></td>
<td></td>
</tr>
<tr>
<td>33. I promise that on Sunday I will write a paper.</td>
<td>This sentence is grammatical because the FocP ‘on Sunday’ does not trigger subject-verb inversion.</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>34. * I promise that on Sunday will I write a paper.</td>
<td>Therefore, the verb and the subject here do not invert.</td>
</tr>
<tr>
<td>35. I promise that never will I tell anyone about this.</td>
<td>This sentence is grammatical because the negative constituent ‘never’ licenses the use of the negative polarity item ‘anyone’.</td>
</tr>
<tr>
<td>36. * I promise that never will I tell someone about this.</td>
<td>This sentence is ungrammatical because the negative constituent ‘never’ licenses the use of the negative polarity item ‘anyone’; yet no negative polarity item is used here.</td>
</tr>
<tr>
<td>37. This is a book which on Sundays I would never read.</td>
<td>This sentence is grammatical because the sequence relativization + topicalization + negative constituent is allowed in English.</td>
</tr>
<tr>
<td>38. * This is a book on Sundays which I would never read.</td>
<td>This sentence ungrammatical because the sequence topicalization + relativization + negative constituent is not allowed in English.</td>
</tr>
<tr>
<td>10. * seem/be asymmetry</td>
<td></td>
</tr>
<tr>
<td>39. John seems to me quite happy with the solution.</td>
<td>This sentence is grammatical because the lexical verb ‘seem’ assigns a theta role to its prepositional phrase (PP) complement ‘to me’.</td>
</tr>
<tr>
<td>40. * John is to me quite happy with the solution.</td>
<td>This sentence is ungrammatical because copula ‘be’ does not assign a theta role to the PP complement ‘to me’.</td>
</tr>
</tbody>
</table>
Table 2. The linguistic patches.

<table>
<thead>
<tr>
<th>Test item</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. You or he is leaving tomorrow.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>2. You or him are leaving tomorrow.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>3. A box or a box of paper is on the floor.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>4. A box or a box of paper are on the floor.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>5. You or I am not working next week.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>6. You or I are not working next week.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>7. A box of paper or two books is on the floor.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>8. A box of paper or two books are on the floor.</td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td></td>
<td>NPs in subject position.</td>
</tr>
<tr>
<td>---</td>
<td>----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>9.</td>
<td>You or I am wrong.</td>
</tr>
<tr>
<td></td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>10.</td>
<td>You or I are wrong.</td>
</tr>
<tr>
<td></td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>11.</td>
<td>Two books or a box of paper is on the floor.</td>
</tr>
<tr>
<td></td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>12.</td>
<td>Two books or a box of paper are on the floor.</td>
</tr>
<tr>
<td></td>
<td>This sentence is a patch because the normal syntactic processes do not license the use of copula ‘be’ with disjoint NPs in subject position.</td>
</tr>
<tr>
<td>13.</td>
<td>Me and her went to a show.</td>
</tr>
<tr>
<td></td>
<td>This sentence conforms to the normal syntactic processes. Yet, prescriptive grammatical rules of English disallows such a construction.</td>
</tr>
<tr>
<td>14.</td>
<td>She and I went to a show.</td>
</tr>
<tr>
<td></td>
<td>This sentence is a linguistic patch because it violates the normal syntactic processes. It is an example of the linguistic patch ‘…and I’ rule.</td>
</tr>
<tr>
<td>15.</td>
<td>For Mary to be the winner and [SCk [IP I the loser] is unfair.</td>
</tr>
<tr>
<td></td>
<td>This sentence is a linguistic patch because it is insensitive to the abstract unpronounced hierarchical constituents. According to the normal syntactic processes, the subject of the small clause ‘I’ should have had the accusative case from the preposition ‘for’ whose copy is in the higher structure [SC]. This is because prepositions only assign</td>
</tr>
</tbody>
</table>
16. For Mary to be the winner and me the loser is unfair. 
   This sentence conforms to the normal syntactic processes. Yet, it is disallowed by prescriptive grammar rules of English.

17. There’s books on the table. 
   This sentence conforms to the normal syntactic processes. Yet, prescriptive rules of English disallow such a construction.

18. There are books on the table. 
   This sentence is a linguistic patch. It is not processed by the normal syntactic processes. Rather, it is enforced by prescriptive grammatical rules of English.

19. There’s a pen and a stamp on the desk. 
   This sentence conforms to the normal syntactic processes.

20. There are a pen and a stamp on the desk. 
   This sentence is a linguistic patch. It is dictated by the prescriptive grammar rules of English.

Table 3. The overall performance of individual ESL participants and NSs on task 1 (TTGJ).

<table>
<thead>
<tr>
<th>#</th>
<th>ESL grp</th>
<th>Correct rating %</th>
<th>#</th>
<th>NS grp</th>
<th>Correct rating %</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>H.A.</td>
<td>60</td>
<td>1</td>
<td>T.V.</td>
<td>73</td>
</tr>
<tr>
<td>2</td>
<td>L.O.</td>
<td>55</td>
<td>2</td>
<td>N.K.</td>
<td>73</td>
</tr>
<tr>
<td>3</td>
<td>R.B.</td>
<td>63</td>
<td>3</td>
<td>N.A.</td>
<td>73</td>
</tr>
<tr>
<td>4</td>
<td>Mi.K.</td>
<td>63</td>
<td>4</td>
<td>A.K.</td>
<td>80</td>
</tr>
<tr>
<td>5</td>
<td>K.</td>
<td>70</td>
<td>5</td>
<td>H.M.</td>
<td>78</td>
</tr>
<tr>
<td>6</td>
<td>E.X.</td>
<td>63</td>
<td>6</td>
<td>P.A.</td>
<td>75</td>
</tr>
<tr>
<td>7</td>
<td>Ma.Kh.</td>
<td>73</td>
<td>7</td>
<td>S.L.</td>
<td>65</td>
</tr>
<tr>
<td>8</td>
<td>J.S.</td>
<td>63</td>
<td>8</td>
<td>K.R.</td>
<td>85</td>
</tr>
<tr>
<td>9</td>
<td>M.M.</td>
<td>58</td>
<td>9</td>
<td>R.J.</td>
<td>78</td>
</tr>
<tr>
<td>10</td>
<td>S.S.</td>
<td>80</td>
<td>10</td>
<td>G.S.</td>
<td>83</td>
</tr>
<tr>
<td>11</td>
<td>W. C.</td>
<td>60</td>
<td>11</td>
<td>A.I.</td>
<td>78</td>
</tr>
<tr>
<td>12</td>
<td>J.B.</td>
<td>68</td>
<td>12</td>
<td>I.</td>
<td>83</td>
</tr>
<tr>
<td>#</td>
<td>ESL grp</td>
<td>Correct preference %</td>
<td>#</td>
<td>NS grp</td>
<td>Correct preference %</td>
</tr>
<tr>
<td>----</td>
<td>---------</td>
<td>----------------------</td>
<td>----</td>
<td>--------</td>
<td>----------------------</td>
</tr>
<tr>
<td>1</td>
<td>H.A.</td>
<td>65</td>
<td>1</td>
<td>T.V.</td>
<td>90</td>
</tr>
<tr>
<td>2</td>
<td>L.O.</td>
<td>55</td>
<td>2</td>
<td>N.K.</td>
<td>95</td>
</tr>
<tr>
<td>3</td>
<td>R.B.</td>
<td>55</td>
<td>3</td>
<td>N.A.</td>
<td>95</td>
</tr>
<tr>
<td>4</td>
<td>Mi.K.</td>
<td>70</td>
<td>4</td>
<td>A.K.</td>
<td>85</td>
</tr>
<tr>
<td>5</td>
<td>K.S.</td>
<td>75</td>
<td>5</td>
<td>H.M.</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>E.X.</td>
<td>75</td>
<td>6</td>
<td>P.A.</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>Ma.Kh.</td>
<td>70</td>
<td>7</td>
<td>S.L.</td>
<td>55</td>
</tr>
<tr>
<td>8</td>
<td>J.S.</td>
<td>65</td>
<td>8</td>
<td>K.R.</td>
<td>90</td>
</tr>
<tr>
<td>9</td>
<td>M.M.</td>
<td>85</td>
<td>9</td>
<td>R.J.</td>
<td>80</td>
</tr>
<tr>
<td>10</td>
<td>S.S.</td>
<td>75</td>
<td>10</td>
<td>G.S.</td>
<td>85</td>
</tr>
<tr>
<td>11</td>
<td>W. C.</td>
<td>75</td>
<td>11</td>
<td>A.I.</td>
<td>100</td>
</tr>
<tr>
<td>12</td>
<td>J.B.</td>
<td>60</td>
<td>12</td>
<td>I.</td>
<td>75</td>
</tr>
<tr>
<td>13</td>
<td>W.I.</td>
<td>60</td>
<td>13</td>
<td>K.N.</td>
<td>85</td>
</tr>
<tr>
<td>14</td>
<td>Ju.B.</td>
<td>45</td>
<td>14</td>
<td>D.E.</td>
<td>85</td>
</tr>
<tr>
<td>15</td>
<td>Y.J.</td>
<td>80</td>
<td>15</td>
<td>B.D.</td>
<td>90</td>
</tr>
<tr>
<td>16</td>
<td>R.F.</td>
<td>85</td>
<td>16</td>
<td>C.W.</td>
<td>50</td>
</tr>
<tr>
<td>17</td>
<td>M.K.</td>
<td>85</td>
<td>18</td>
<td>M.L.</td>
<td>45</td>
</tr>
<tr>
<td>19</td>
<td>H.F.</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4. The overall performance of individual ESL participants and NSs on task 3 (the preference task)
Table 5. The performance of ESL group and the NS group on task 2 (the correction task)

<table>
<thead>
<tr>
<th>Syntactic categories</th>
<th>Impossible sentences</th>
<th>% of successful corrections among ESL grp</th>
<th>% of successful corrections among NS grp</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subjacency</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Wh-islands</td>
<td>1. * Where did they wonder whether she would leave the key?</td>
<td>40%</td>
<td>81%</td>
</tr>
<tr>
<td>2. NP complement</td>
<td>2. * Where did the fact that Morag put the ice cream make you angry?</td>
<td>20%</td>
<td>69%</td>
</tr>
<tr>
<td>islands</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. CP adjunct</td>
<td>3. * How did they leave the pub by car before the game finished?</td>
<td>15%</td>
<td>19%</td>
</tr>
<tr>
<td>islands</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. The ECP Principle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. Superiority</td>
<td>4. * What did he wonder who said?</td>
<td>30%</td>
<td>56%</td>
</tr>
<tr>
<td>b. That-trace</td>
<td>5. * Who did you wonder whether called?</td>
<td>35%</td>
<td>69%</td>
</tr>
<tr>
<td>effect</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>c. NP movement</td>
<td>6. * Who do you think that will invite Mary?</td>
<td>30%</td>
<td>100%</td>
</tr>
<tr>
<td>5. The Binding</td>
<td>7. * John seems that it is likely t to go.</td>
<td>30%</td>
<td>44%</td>
</tr>
<tr>
<td>Principle</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. The AP-</td>
<td>8. * Julie believes John to expect pictures of herself to be on sale next week.</td>
<td>20%</td>
<td>31%</td>
</tr>
<tr>
<td>internal subject</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hypothesis</td>
<td>9. * John considers Mary to be too proud of himself.</td>
<td>30%</td>
<td>81%</td>
</tr>
<tr>
<td>6. Case Theory</td>
<td>10. * It was thought Mary arriving late.</td>
<td>70%</td>
<td>81%</td>
</tr>
<tr>
<td>11. * John seems that he likes Mary.</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a. The AP-</td>
<td>12. * There were very three students sick.</td>
<td>30%</td>
<td>13%</td>
</tr>
<tr>
<td>internal subject</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>hypothesis</td>
<td>13. * To get not arrested for murder is a</td>
<td>25%</td>
<td>31%</td>
</tr>
<tr>
<td>7. Finite/non-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>finite asymmetries</td>
<td>miracle.</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-------------------</td>
<td>----------</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>8. Theta-Theory a. Copula be vs. seem</td>
<td>14. * John is to me quite happy with the solution.</td>
<td>45%</td>
<td>44%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Head-Movement Constraint (HMC)</td>
<td>15. * I promise that never will on Sundays I drink.</td>
<td>95%</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td>16. * Why did on Sunday they invite Tom?</td>
<td>65%</td>
<td>88%</td>
</tr>
<tr>
<td>a. Whether/if asymmetry</td>
<td>17. * I wonder if or not they will come.</td>
<td>65%</td>
<td>63%</td>
</tr>
<tr>
<td>10. Negative constituents vs. focus items as triggers of inversion and negative polarity items</td>
<td>18. * I promise that on Sunday will I write a paper.</td>
<td>75%</td>
<td>56%</td>
</tr>
<tr>
<td>a. Negative constituent vs. focus items as triggers of inversion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>b. Negative constituents as triggers of negative polarity items</td>
<td>19. * I promise that never will I tell someone about this.</td>
<td>5%</td>
<td>25%</td>
</tr>
<tr>
<td>c. Relativization + topicalization + negative constituent</td>
<td>20. * This is a book on Sundays which I would never read.</td>
<td>40%</td>
<td>56%</td>
</tr>
</tbody>
</table>
Table 6. The performance of the ESL group on linguistic patches

<table>
<thead>
<tr>
<th>Linguistic viruses/patches</th>
<th>Participants whose performance is in accordance with viruses/patches (n = 20)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. You or he is leaving tomorrow.</td>
<td>90%</td>
</tr>
<tr>
<td>2. You or him are leaving tomorrow.</td>
<td>90%</td>
</tr>
<tr>
<td>3. Me and her went to a show.</td>
<td>80%</td>
</tr>
<tr>
<td>4. She and I went to a show.</td>
<td>100%</td>
</tr>
<tr>
<td>5. There’s books on the table.</td>
<td>80%</td>
</tr>
<tr>
<td>6. There are books on the table.</td>
<td>100%</td>
</tr>
<tr>
<td>7. A book or a box of paper is on the floor.</td>
<td>95%</td>
</tr>
<tr>
<td>8. A book or box of paper are on the floor.</td>
<td>95%</td>
</tr>
<tr>
<td>9. You or I am not working next week.</td>
<td>90%</td>
</tr>
<tr>
<td>10. You or I are not working next week.</td>
<td>85%</td>
</tr>
<tr>
<td>11. For Mary to be the winner and I the loser is unfair.</td>
<td>55%</td>
</tr>
<tr>
<td>12. For Mary to be the winner and me the loser is unfair.</td>
<td>55%</td>
</tr>
<tr>
<td>13. There’s a pen and a stamp on the desk.</td>
<td>45%</td>
</tr>
<tr>
<td>14. There are a pen and a stamp on the desk.</td>
<td>70%</td>
</tr>
<tr>
<td>15. A box of paper or two books is on the floor.</td>
<td>95%</td>
</tr>
<tr>
<td>16. A box of paper or two books are on the floor.</td>
<td>95%</td>
</tr>
<tr>
<td>17. You or I am wrong</td>
<td>80%</td>
</tr>
<tr>
<td>18. You or I are wrong.</td>
<td>85%</td>
</tr>
<tr>
<td>19. Two books or a box of paper is on the floor.</td>
<td>85%</td>
</tr>
<tr>
<td>20. Two books or a box of paper are on the floor.</td>
<td>85%</td>
</tr>
</tbody>
</table>

M = 83 %
Appendix B

Biographical Questionnaire

NAME___________________________

1. What is the first language that you learned? ______________

2. What languages do you speak at your current residence in Canada? Circle only ONE option.
   a. Always English   b. Mostly English       c. Always my first language       d. Mostly my first language
   e. Neither my first language nor English

3. What language do you speak outside your current residence in Canada? Circle only ONE option.
   a. Always English   b. Mostly English       c. Always my first language       d. Mostly my first language
   e. Neither my first language nor English

4. In addition to your first language and English, do you speak any other language?

   No ____       Yes ___

   If yes, please say which language:____________

   How well do you know this language?

   1 2 3 4 5
   Not at all  completely fluent

5. If you lived anywhere else between leaving your home country and arriving in Canada, please state where and for how long:

   Country:_______________                          ___ years  ___ months

6. How long have you been living in Canada?       ___ years ___ months
7. How long have you studied English in an English/ESL class?

   In Canada (any ESL school other than the current one):
   __ years __ months  Number of hours per day ______

   In your home country ______ years ____ months  Number of hours per day ______

8. Which of these did you attend in your home country?

   School  College/University
   Yes _____  Yes _____
   No _____  No _____
   Number of years _____  Number of years _____

9. What is your major/specialization?

10. Did you attend any other schools in Canada before you joined the ESL program at the University of Toronto?

    Yes _____
    No _____
    Number of months /years _____

11. How long have you been studying at the ESL program of the University of Toronto? _____ years _____ months

12. Which, if any, of the following English proficiency tests have you taken?

    | Test         | Year (e.g., 2010) | Score |
    |--------------|-------------------|-------|
    | TOEFL        |                   | _____ |
    | MELAB        |                   | _____ |
    | ELDA/COPE    |                   | _____ |
    | TSE          |                   | _____ |
    | IELTS        |                   | _____ |
    | CAEL         |                   | _____ |
    | Other:       |                   | _____ |

13. Which age group do you fall into?

    18-29  30-39  40-49  50 or above
14. Gender:  
   Female _____  Male _____

15. Throughout your study of English, have you received any instruction in grammar?  
   Yes ____  
   No _____  
   Number of months/years ____