Arabic and Chinese Learners' Production, Perception and Processing of Past Tense and Verbal Agreement Morphology in L2 English

Walid Kahoul

A thesis submitted to the

School of English Literature, Language and Linguistics

for the degree of

Doctor of Philosophy

At

Newcastle University

May, 2014

Abstract

This thesis is an attempt to locate the source of the difficulties experienced by adult second language (L2) learners of English in the production of past tense and verbal agreement morphology, observed as an alternation between the inflected and stem forms of verbs in contexts where only the former is accepted in target grammars. To this end, current competing syntactic, non-syntactic and phonological accounts of the phenomenon are tested against production, perception and processing data.

Production was tested by a sentence elicited imitation task, which comprised of 50 aural sentences creating obligatory contexts for the properties under study. Participants were asked to repeat the sentences one by one and their response was recorded. Perception and processing were tested by a computerised picture-choice task, which consisted of 88 picture-sentence trials. Participants were asked to choose one picture, the choice of which depended on their perception of verbal morphology. Picture choice response, reaction times and eye movements were recorded in this task. Thirty-seven L1 speakers of Chinese and thirty-four L1 speakers of Arabic, who were matched in L2 proficiency at low, mid and high levels, in addition to a control group of ten native speakers of English participated in the study.

The results of the production and perception studies similarly showed that while Chinese participants produced and perceived the morphology variably at all levels, Arab participants did so only at low and mid levels, overcoming variability at the highest proficiency level. Neither production nor perception data demonstrated phonological effects. Results from the processing study revealed that both language groups processed the morphology similarly at low and mid levels but they differed at the high level with only Arab participants' data showing evidence for developing automatic competence. These findings strongly suggest that morphological variability is caused by absence of syntactic representations which are built up incrementally with rising proficiency supporting structure building accounts of L2 acquisition.

Acknowledgements

I would like to express my deep gratitude to my supervisor, Professor Martha Young-Scholten, for her great support and invaluable advice. Without Prof. Young-Scholten's academic guidance, this thesis would not have been possible. My heartfelt gratitude and appreciation is extended to my co-supervisor, Dr. Clare Wright, whose encouragement and support to me has always been inspirational. Dr. Wright's thought-provoking comments have benefited this research immensely. I would like to gratefully and sincerely thank Dr. Nick Riches, who replaced Dr. Wright when she moved to Reading University, for his insightful comments during the writing up of this thesis. Sincere thanks are due to Professor Roger Hawkins and Dr. Geoffrey Poole for kindly being my examiners.

I am extremely grateful to Dr. Chris Petkov and Dr. Quoc Vuong, who made the eye-tracking study possible by allowing me unrestricted access to the eye-tracking laboratory at the Institute of Neuroscience at Newcastle University and providing guidance and assistance in setting up the experiment. I owe a great debt of gratitude to Dr. Shahin Mojarad, who did the MATLAB programming for the eye-tracking experiment. My grateful thanks go to Dr. Sami Ramadan for helping me with the experiments and Dr. Amer Alkafri for the enjoyable discussions we had.

I am also grateful to the Chinese and Arab informants who generously participated in the studies without anything in return.

Last but not least, I would like to thank Rima, my wife, for her patience and tolerance over the last four years. I wish to acknowledge my late father, mother, sister and brothers as being the source of inspiration in my studies. I would particularly thank Marwan, my brother, whose support was inestimable.

To the martyrs of the Dignity Revolution in Syria

Table of Contents

List of Figures

List of Tables

Chapter 1. Introduction

1.1 The Research Phenomenon

Learning a second language (L2) is not an easy feat. To express an idea appropriately and get a linguistic message through to other interlocutors, one has to choose the lexical items that are suitable for the context and combine them together according to the grammar of the language. Classroom teachers would agree that learning the grammar of an L2 and incorporating this knowledge in real time production are areas of difficulty for language learners; even high proficiency learners might sometimes make grammatical errors when they speak the L2. One area of specific difficulty for L2 learners (L2ers) is the production of free or bound morphemes related to functional categories such as complementisers, inflections and determiners. The English regular past tense marker *–ed* (as in 'he *played* football') and the verbal agreement marker *–s* (as in 'she *goes* to school') are examples of bound morphemes and these are the focus of the present study.

It is widely observed that L2ers in spontaneous production sometimes drop functional morphemes where these are obligatorily required by the target language, which surfaces as an inconsistent use of these morphological items (e.g., Haznedar and Schwartz, 1997; Lardiere, 1998a, 1998b; Ionin and Wexler, 2002). This target-deviant performance is known as morphological variability and it has been attested in the acquisition of various L2s by learners from different first language (L1) backgrounds (see e.g., Vainikka & Young-Scholten 1994, 1996 for the acquisition of L2 German by L1 speakers of Korean, Turkish, Spanish or Italian; Prévost, 2003 for the acquisition of L2 French by L1 speakers of English; Campos Dintrans, 2011 for the acquisition of L2 English by L1 speakers of Chinese, Japanese or Spanish; Oldenkamp, 2013 for the acquisition of L2 Dutch by L1 speakers of Turkish, Arabic or Chinese). Moreover, this inconsistent use of functional morphemes was found to be experienced not only by learners at low levels of L2 proficiency (Vainikka & Young-Scholten, 1994, 1996b) but also by more proficient learners (Hawkins and Liszka, 2003) and even sometimes by learners who are apparently at the end state of their acquisition (Lardiere, 2007, 2013). The phenomenon is robustly observed but its source and the reason for its persistence are hotly debated.

Morphological variability in adult second language acquisition (SLA) is the phenomenon addressed by this thesis. The framework adopted in this endeavour is generative grammar (Chomsky, 1959, 1965, 1981, 1995, 2000). A core assumption in this framework is that an abstract subconscious linguistic system underlies language use and native grammars are constrained by a set of innate principles and parameters called Universal Grammar (UG). A great deal of studies in both L1 and L2 acquisition has been conducted within this framework and in the past few decades, it has proven to be a "successful approach to understanding the mechanisms which underlie the human ability to build mental grammars" (Hawkins, 2001: 1).

The acquisition of functional morphology and their underlying syntactic representations (morpho-syntax) by L2ers is a topic of central investigation in current generative SLA research (Hawkins, 2009; Ionin, 2013). While it is not disputed that L2ers have difficulties with the acquisition of morphosyntactic properties as manifested, for example, by the phenomenon under investigation in this thesis (i.e., morphological variability), the source of these difficulties is, however, hotly debated. The availability of UG and the role of the learners' native language in SLA have been core issues in all generative attempts to explain the target-deviant performance.

Previous research investigating morphological variability has relied mainly on production, and to far less extent, perception data. Although a growing body of research in the field of psycholinguistics has started to focus on how L2ers process functional morphemes in real time comprehension (see e.g., Clahsen, Felser, Neubauer, Sato and Silva, 2010, for a review), little use has been made of its findings to resolve the debates on the source of morphological variability. Assuming that the same syntactic representations are involved in language production, perception and processing, it is held in this thesis that the study of any of these modalities should inform us about the syntactic representations underlying them and thus helps uncover the source of morphological variability.

1.2 Method and the Research Question

This thesis investigates the phenomenon of variability in the use of English past tense and verbal agreement morphology in SLA in the aim of locating its source and the reason for its persistence. Different from previous research on the phenomenon, which relied mainly on production data, this thesis benefits from production, perception and processing data to resolve the theoretical controversies on the source of the problem.

To investigate the role of the mother tongue in morphological variability, the study of this thesis includes native speakers of Arabic or Mandarin Chinese. The rationale for including speakers from these L1 backgrounds is the similarities and dissimilarities these languages bear to each other, in one hand, and to English (the L2 target), on the other, with regard to the properties under investigation. While Arabic has syntactic features for past tense and verbal agreement and, at the syllable level, allows the structure in which the English inflection is sometimes realized (i.e., consonant clusters), Mandarin Chinese lacks syntactic features for these properties and disallows consonant clusters. Moreover, both Arabic and Chinese lack the prosodic structure required for accommodating the English inflection (i.e., adjunction to prosodic word). Therefore, in a carefully designed study which controls for the phonological structure to tease apart different potential sources of L1 transfer, data from speakers of these languages will provide a good testing ground to many proposals on the role of the L1.

Furthermore, in order to obtain a better understanding of morphological variability, this thesis adopts a cross-sectional design including L2ers from a range of proficiency levels: Low, Mid and High. Since previous research has shown that this phenomenon is experienced by learners at different levels of linguistic development as mentioned above, it is conceivable to believe that investigating the phenomenon in one specific proficiency level might not reveal all the facts about it and, conversely, studying it at more than one proficiency level would give wider insights.

As the language experience varies from one learner to another and the study includes participants from two backgrounds, it is an essential procedure to implement a means to match and compare participants with each others. The means followed in this thesis is a proficiency measure. Particularly, Unsworth's (2005, 2008) Age-Sensitive Composite Proficiency Score (ASCOPS) is adopted. ASCOPS was performed on semi-spontaneous data collected from the participants in this study using a picture description task. The rationale for using this measure rather than any other means is that it allows gauging the participants' L2 proficiency irrespective of their performance on the specific properties under investigation (i.e., past tense and verbal agreement), which should not be a determinant factor in the participants' proficiency to avoid a confound in this study.

The experimental method used to investigate the participants' production of past tense and verbal agreement is a sentence elicited imitation task. The stimuli in this task are designed in such a way as to allow testing the production of the relevant morphological items in different phonological contexts. That is, it includes both verbs that create a word-final consonant cluster structure when the *-ed* or *-s* inflections are added (e.g., walked /kt/; travelled /ld/; walks /ks/; travels /lz/) and verbs that do not create a consonant cluster structure when the same inflections are added (e.g., played /eid/; plays /eiz/). By doing this, the data collected using this task will provide good testing grounds for the phonological accounts of the phenomenon of morphological variability. It will also allow teasing apart the possible syntactic and phonological sources of L1 transfer.

The experimental method used to investigate the participants' perception and processing of past tense and verbal agreement morphology is a computerised picture-choice task supplemented by reaction time and eye-tracking response measures. Similar to the production task, this task includes verbs from different phonological structures to investigate any phonological effects on the perception and processing of the inflection.

Based on the goals defined in this section, this thesis addresses the following research question:

> *What is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?*

1.3 The Outline of this Thesis

Chapter two provides the theoretical background of this research. It starts with reviewing core concepts in the generative grammar framework, which is the approach adopted in this thesis. It then discusses some issues related to SLA of morphosyntax and addresses the research phenomenon, i.e., morphological variability, and its theoretical accounts in generative SLA. After that, the chapter reviews some findings on L2 processing of morphosyntax and discuses possible factors affecting it. Chapter three focuses on the difficulty L2ers have with the use of past tense and verbal agreement in particularly L2 English and, thus, reviews previous research on the production, perception and processing of these properties. The same chapter also describes the characteristics of the properties under study in the research participants' L1s (Arabic and Chinese) and the target L2 (English) and provides the linguistic assumptions at the syntactic and phonological levels. The participants in this research and their biographical information as well as the measure used to assess their L2 proficiency are described in Chapter four. The results of the proficiency measure are reported in the same chapter. Chapters five, six and seven present the production, perception and processing studies, respectively. Each of these chapters describes the methodology used in the study and reports on its results. Chapter eight brings together and discusses the findings of the three studies in relation to the literature discussed in chapters two and three and details the main conclusions

Chapter 2. Second Language Acquisition of Morphosyntax

2.1 Introduction

The aim of this chapter is to introduce the theoretical framework in the light of which this research on adult SLA of some morphosyntactic properties, specifically past and verbal agreement, should be understood. The chapter starts with outlining the main premises of the approach to language acquisition this thesis adopts, that is, generative grammar. Then, two classifications on syntactic categories and linguistic features relevant to the topic of acquisition of morphosyntax will be introduced. This will be followed by a sketch of the relationship between morphology and syntax as maintained by research in generative SLA. After that, morphological variability, the particular phenomenon under investigation in this thesis, as documented in empirical research along with its proposed theoretical accounts will be presented and discussed. Finally, shifting to the psycholinguistic field of study, how L2 processing research findings can serve resolve theoretical controversies in SLA is discussed followed by a review of some findings of studies on L2 processing of morphosyntax.

2.2 The Theoretical Background: Generative Grammar

The linguistic theory within the generative grammar framework has formed the theoretical background for a good deal of research conducted in the fields of L1 and L2 acquisition. It is worth starting here by outlining some of the main premises of the theory of generative grammar as they pertain inextricably to most of the hypotheses and ideas presented, tested and discussed in this thesis. This section touches particularly on the generative suppositions that children are born with an innate ability to learn language. It then introduces two different classifications made in generative syntactic theory to syntactic categories (i.e., functional and lexical) and formal linguistic features (i.e., interpretable and uninterpretable), which are, as will be seen in a later section, relevant to the investigation of the phenomenon of morphological variability in SLA.

Within the generative grammar framework, (Chomsky 1959, 1965 and much subsequent work), it is assumed that there is an abstract subconscious linguistic system underlying language production and comprehension. This proposal makes a distinction between the language we produce and comprehend, and the tacit knowledge we have of that language. In fact, this distinction was drawn by Chomsky in the 1960s and he named the former 'performance' and the latter 'competence'. In Chomsky's words, these are defined as that performance is "the actual language use in concrete situations" and competence is "the speaker-hearer's knowledge of his language" (Chomsky, 1965: 4). In language acquisition research, the nature of the linguistic competence and how it develops over time have been the focus of much empirical and theoretical investigation. As we will see latter, this thesis attempts to examine the nature of the linguistic competence of some (specifically second) language learners by inspecting their linguistic performance.

Explaining how children can acquire a complex linguistic competence effortlessly and in a relatively short time despite the meagre input they are exposed to, which is usually referred to as the learnability problem, has been the main task of generative grammar since its beginning in the 1950s (Hornstein, Nunes and Grohmann, 2005: 2). The innateness of the language learning ability is posited as a solution as thus. It is maintained that the starting point of first language acquisition (FLA) is a supply of innate linguistic knowledge provided by the so called language faculty (Chomsky, 1965, 1972). This is to say that all children are born with a linguistic endowment to help them parse and acquire the language they are exposed to. This inborn knowledge is called Universal Grammar (UG) and it is believed to comprise all core grammars relevant to all natural languages. Therefore, as children are exposed to the primary linguistic data (PLD) in their environment, the task of UG is purported to guide and facilitate acquisition of the ambient language.

Indeed, all typically developing children by the age of five reach a stage where they have acquired the core grammars of their language and this happens with relative ease despite the complexity of the linguistic system of that language (Guasti, 2002). Moreover, although children receive a linguistic input that might be imperfect, i.e., "a good deal of normal speech consists of false starts, disconnected phrases, and other deviations from idealised competence" (Chomsky, 1972: 158), and their linguistic experience is based solely on positive evidence (i.e., only constructions that are acceptable in their language), they acquire the language successfully and they come to know not only what the grammatical system of their language allows but also what it disallows (see, e.g., Hornstein and Lightfoot, 1981). Therefore, generative linguists hold that the linguistic input that children have access to in their environment is insufficient to explain the grammatical competence they attain. It is rather believed that children's rapid, effortless and successful acquisition of a highly complex linguistic system cannot come to fruition without some kind of prior knowledge, presumably UG (Chomsky, 1965).[1]

The nature of UG is another important issue in generative grammar. Within the Principles and Parameters approach (Chomsky, 1981), UG is conceived as a set of invariant principles that govern all natural languages (i.e., linguistic aspects that are universal to all languages) and a number of parameters that finitely vary in their values (or settings) across languages (i.e., language-particular aspects). While the principles account for the commonalities among languages, the parameters are posited to capture the differences across languages. A core implication this model has on language acquisition is that the learner's task, at the grammatical level, is restricted to acquiring the parameters' settings of the target language based on the evidence available through the speech input.

In a more recent generative approach, namely Minimalism (Chomsky 1993, 1995, 2000), it is hypothesised that in addition to the universal principles, UG provides a universal inventory of formal linguistic features and computational operations (Merge, Agree and Move). The inventory of features consists of all syntactic, semantic and phonological features that exist in all human languages (e.g., person, number, case, finiteness, tense, gender, etc). Cross-linguistic variation, which was captured by the idea of different parameter settings in the Principles and Parameters approach, is now assumed to be due to the different selections each language makes from this inventory of features. In this model, part of the child's acquisition task is speculated to be selecting a subset of this inventory based on the PLD received from the ambient environment (Chomsky, 2001).

A core issue within this framework that is relevant to the topic of this thesis is a classification made to syntactic categories and formal features. This is dealt with hereafter.

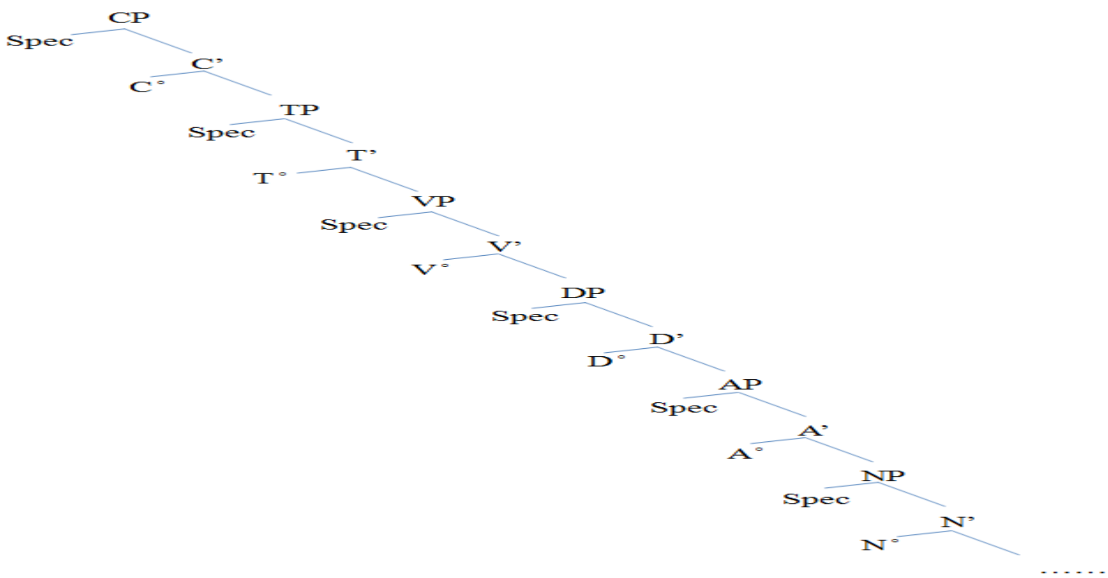
[1] The space limit here precludes discussing the full range of arguments on the innateness hypothesis; see Antony and Hornstein (2002) for further discussion.

2.2.1 *Syntactic Categories and Formal Features*

In Minimalism (Chomsky, 1995, 2000, 2001), lexical items are believed to be formed from bundles of formal, semantic and phonological features and these provide the input to a subset of the computational system, namely the syntax, which transforms them by a series of computational procedures into a syntactic structure of a given linguistic expression. The resulting syntactic structure serves as input into two other representational systems, which are the Phonetic Form (PF; i.e., a representation of the phonetic form of language, dictating how it is pronounced) and the Logical Form (LF; a representation of the meaning of language) (Chomsky, 1993). With this conception in mind, I shall introduce hereafter two classifications made to syntactic categories and formal features and clarify how these classifications are relevant to this thesis.

In generative grammar, syntactic structures are argued to be binary-branching trees (Kayne, 1984) consisting of hierarchically organised phrase projections (e.g., Noun phrase (NP), adjective phrase (AP), determiner phrase (DP) verb phrase (VP), tense phrase (TP) and complementiser phrase (CP)) as shown in Figure 2.1.

Figure 2.1: Illustration of phrase structure in Generative Grammar



Generative grammarians make a distinction in the syntactic structure between lexical and functional categories (Chomsky, 1986; Abney, 1987). Basically, lexical categories are headed by elements that carry meaning about a linguistic expression whereas functional categories are headed by elements that perform a grammatical function in a linguistic expression (Radford, 2004). For example, NP, AP and VP (headed by nouns, adjectives and verbs respectively) are classified as lexical categories and on the other hand DP, TP, and CP (headed by determiners, tense markers and complementisers respectively) are classified as functional categories.

Under Minimalism, the heads of functional categories consist of features that need to be checked by the constituent features of the lexical items which enter the derivation. One issue that is of an interest to us here is a classification that is made to these features according to their relevance to the semantic interpretation as interpretable or uninterpretable features (Chomsky, 2001). While the former are readable at the LF (i.e., they are relevant to the semantic interpretation of the expression), the latter are not (examples below). However, although uninterpretable features are void of semantic content, they are necessary as they perform functions relevant to the grammaticality of the expression and can be readable at PF (i.e., they can have phonological manifestations). Radford (2004: 287-288) gives an illustrative example for these types of features: pronouns in English have a Number, Person and Case features. Number and Person are interpretable features as there is clearly a difference in meaning between a pronoun such as *I*, which refers to a first-person-singular subject, and another such as *they*, which refers to a third-person-plural subject. By contrast, Case is uninterpretable. A comparison of a, b and c in Example 2.1 shows that there is no difference in meaning among *they*, *them* and *their* which are marked with different cases (i.e., nominative, accusative and genitive, respectively).

(2.1) (a) It is said [*they* were arrested]
(b) He expected [*them* to be arrested]
(c) He was shocked at [*their* being arrested]
(from Radford, 2004: 288)

However, although there is no difference in meaning among the subject pronouns in the bracketed clauses in the three sentences of Example 2.1, they have different phonological forms reflecting their different grammatical functions within the relevant phrase and any attempt to replace one form by another would result in an ungrammatical sentence (e.g., *He expected *they* to be arrested). Therefore, uninterpretable features do not have a semantic meaning but they might have a phonological realization and they perform a grammatical function in a linguistic expression.

I shall now show how these divisions are relevant to this research.

As stated above, in Minimalism it is assumed that differences among languages arise from the different selections each language makes from the inventory of formal features. Specifically, features on functional heads (e.g., D, T, C) are thought to be the locus of parametric variations across languages (Chomsky 1993, 1995, 2001). This has an important implication for research in the field of SLA (and thus the research at hand); that is, a certain feature might be instantiated in certain languages but not others, and L2ers are faced with the task of acquiring new features that are not present in their L1s. Indeed, whether L2ers can acquire new features not present in their native languages is a research question that has been under investigation for the past two decades in generative SLA. Ionin (2013: 505) points out that:

> The central question in generative approaches to the SLA of morphosyntax is whether L2 learners (in particular adult L2 learners) are capable of constructing a target-like syntactic representation, especially in those domains where the learners' native language and their target language differ.

The interpretable/uninterpretable distinction between formal features is relevant here. Some researchers (Hawkins and Chan, 1997; Hawkins and Liszka, 2003) argue that uninterpretable features are subject to a critical period and, thus, post-puberty L2ers cannot acquire such features if they are not instantiated in their L1 (see 2.3.3.3 below).

For this thesis, the formal features that are of most interest are component features of the English functional category TP and specifically those responsible for ensuring past tense and verbal agreement[2] are expressed on verbs. As we will see later (Chapter three, section 3.3), these properties are associated with uninterpretable features which are also

[2] Under some syntactic accounts (e.g., Split-INFL hypothesis (Pollock, 1989); Agr-based theory of clause structure (Chomsky, 1991, 1993)), features associated with verbal agreement are assumed to project in a separate category, i.e., AgrP. However, this idea has been abandoned in Chomsky's subsequent work (1995, 1998, 2001) and a single projection (i.e., TP) is adopted instead.

parameterised in the sense that they exist in certain languages but not others. This thesis explores the state of these features in SLA through examining the productive and receptive use of their phonological exponents by L2ers who have or do not have them in their L1s. This will be in the aim of locating the source of morphological variability experienced by adult L2ers.

2.3 Second Language Acquisition of Morphosyntax

The nature of the L2ers' mental representations (interlanguage) and how UG and the native language linguistic system are involved in building or constraining them are issues that have dominated the field of generative SLA since the 1970s (see e.g., Hawkins, 2001 and White, 2003b). One motivation for these enquiries is the observed inaccuracies in the use of certain morphosyntactic properties in the speech of L2ers at different points of linguistic development. Functional morphology such as past and verbal agreement marking on verbs is one domain in which such inaccuracies have been observed (Lardiere, 2007). These inaccuracies are exhibited in the L2ers's speech as an inconsistent use of functional morphemes in obligatory contexts, diverging from the expected native speaker patterns. There is no consensus in the field on the source of this inconsistency and the interpretation to be given to this phenomenon is hotly debated. Does the inconsistent use of the functional morpheme reflect non-target like underlying syntactic representations? Or do L2ers have target-like syntactic representations and the source of the observed inconsistencies is non-syntactic? As will be detailed below, answers to these questions have been captured within hypotheses on the nature of the mental representations at the outset of acquisition and how they develop afterwards. The availability of UG as well as the role that the L2ers' native language plays in SLA have been core issues in such hypotheses.

Section 2.3.1 presents an overview of the research phenomenon under investigation; that is, variability in the use of functional morphemes, particularly verbal tense and agreement inflectional morphology. Then, section 2.3.2 sketches views adopted by generative L2 researchers on the relationship between the surface morphology and underlying syntax followed in section 2.3.3 by the main hypotheses on the source of morphological variability.

2.3.1 Morphological Variability: The Phenomenon

A considerable body of research in the field of SLA has focused on the acquisition of functional morphology, specifically tense and agreement marking on verbs. A robust phenomenon that has been observed is that L2ers in spontaneous production sometimes drop the verbal inflections where they are obligatorily required by the target language; that is, the morpheme is used inconsistently as it is sometimes supplied and sometimes dropped. This inconsistency, or variability as we will call it from now onwards, seems a characteristic of the speech of child and adult L2 speakers as found, for example, by the early morpheme order studies on L2 English (e.g., Dulay and Burt, 1973, 1974; Bailey, Madden and Krashen, 1974; Larsen-Freeman, 1975; Krashen, 1977; Andersen, 1978; Makino, 1980) as well as much subsequent research on a variety of L2s, including, but not limited to, French, German and English (e.g., Lardiere 1998a, b; Prévost and White, 2000a, b). To put our discussion on concrete footing, the following utterances in (2.2), (2.3) and (2.4) are presented as examples of speech collected from L2ers of English, French, German respectively (source is provided alongside each example):

(2.2) L2 English
a. and then he # he gained his sight (Patty)
b. yeah, Saul *gain* his sight.
(Lardiere, 2003: 178)

(2.3) L2 French
a. i demande (Abdelmalek)
I ask-1/2/3S[3]
b. pas *demander* les papiers
not ask-INF the papers
(Prévost and White, 2000a: 210)

(2.4) L2 German
a. ich studiere nicht (Zita)
I study-1S[4] not
b. ich *studieren* in Porto
I study-INF[5] in Porto
(Prévost and White, 2000a: 210)

---

[3] Here, the agreement is appropriate but the 1st, 2nd and 3rd person singular subject-verb agreement markers in French are homophonous in form.
[4] '1S' stands for 1st person singular subject-verb agreement
[5] 'INF' stands for infinitive

In these examples, tense and/or subject-verb agreement inflections are dropped from the verbs in (2.2b), (2.3b) and (2.4b) and bare or infinitive forms are supplied instead. The appropriate morphological markers are supplied on the verbs in (2.2a), (2.3a) and (2.4a). These pairs of sentences were uttered by the same learners and during the same interviews.

The rates of inflection suppliance in the speech of these learners who produced (2.2), (2.3) and (2.4) (as well as in the speech of many other L2 learners reported in other studies) reveal that these alternations form a robust observation and they cannot be temporary slips of the tongue or mistakes. Prévost and White (2000a) found that the Arab learner, Abdelmalek (an L2er of French) supplied the appropriate inflection in around 75% of obligatory contexts and the Portuguese learner, Zita, (an L2er of German) did so in around 70% of obligatory contexts as attested in data collected through interviews conducted in a monthly basis over a period of three years with Abdelmalek and two years with Zita. Lardiere (2003) also reported that the production rate of past tense regular inflection was less than 35% in the speech of her informant. Therefore, this target-deviant performance is a robustly-observed characteristic of the speech of many L2ers (for reviews, see Hawkins 2009 and Ionin 2013).

One finding that has arisen from research investigating this phenomenon in adult L2ers' speech is that it might be a characteristic of the speech not only of learners at developmental stages of acquisition, e.g., Vainikka & Young-Scholten (1994, 1996), Goad, White and Steele (2003) and Prévost (2003) among other studies, but also of learners at a stage where no more development is likely to occur, e.g., Lardiere (2007). In Goad et al.'s study, the informants are 12 L2ers of English assessed to be at a high-intermediate low-advanced proficiency range. All of these learners are reported to have produced past and agreement verbal morphology inconsistently. Prévost's study participants are 21 L2ers of French placed at proficiency levels ranging from beginners to low intermediate. Most of these learners also alternated between the use of inflected and non-inflected verbs to a varying degrees that are proportionate with their proficiency levels. Vainikka & Young-Scholten (1994, 1996) studied the speech of 28 Korean, Turkish, Spanish or Italian untutored L2ers of German, who had resided in Germany for a period ranging between 1;5 to 25 years. These learners were placed at different proficiency levels corresponding to syntactic stages. Twenty of them (those who were placed at the two lower syntactic stages) were found highly variable in their production of verbal morphology. This variability in the speech of such proficiency level learners is apparently an aspect of a developmental stage and it could be assumed that these learners would overcome it with more exposure to the target language. However, a longitudinal case study by Lardiere (2007) raises the possibility that morphological variability might persist into a stage at which no more development is expected. Lardiere studied the spontaneous production of a Chinese-speaking L2er of English called Patty. According to the author, Patty was at an end-state of L2 English. Despite this, a high rate of omission of verbal inflections was observed in Patty's speech (a review of this study is presented in Chapter three, section 3.2). Hence, the emerging picture from SLA research exposes morphological variability not only as a developmental phenomenon that all learners might experience, but also as a persistent problem in the speech of some learners.

Another finding that has arisen from L2 research on morphological variability is that regardless of their L1 background, L2ers pass through a stage at which they will variably produce verbal inflections but the L1 might be a factor that determines whether this variability would persist into advanced stages of proficiency or not. The learners who are reported to have used the L2 verbal inflection inconsistently are native speakers of a range of languages that vary in richness in verb morphological paradigms, with no apparent effect on the presence/absence of this linguistic behavior at early stages of SLA. The studies mentioned so far in this section support this finding. The informants included L1 speakers of Arabic or Portuguese as in White and Prévost (2000a), Chinese as in Goad et al. (2003), English as in Prévost (2003) and Korean, Turkish, Spanish or Italian as in Vainikka & Young-Scholten (1994, 1996). All of these learners were reported to have experienced variability in using the L2 Inflection. Portuguese, Arabic, Russian, Turkish, Italian and Spanish are morphologically rich languages, whereas Chinese and Korean are impoverished in this regard and English falls in between. This indicates that whether the learner's L1 has rich or impoverished morphological paradigm does not change their fate in passing through a stage of morphological variability in SLA.

Nevertheless, while there is ample evidence showing that morphological variability might affect all L2 learners regardless of their L1s, recent research has revealed that the L1 of learners could possibly be a factor that determines how persistent this variability might be. Two studies on the acquisition of L2 English that will be reviewed in Chapter three (3.2.1) raise this possibility. These are Lardiere (2007) and White (2003). While the former studied a Chinese-speaking L2er of English, the latter studied a Turkish-speaking L2er of English. Both studies claimed that their informants were at an end state of SLA. As we shall see in more detail in section 3.2.1, while Lardiere reported high rates of omission of the past and agreement inflections in the speech of her informant, White found that the rate of omission of the same inflections in her informant's speech was considerably lower. Although the source of the difference between the two informants is not clear, their performance might be indicative of a possible role for the L1 as they speak two typologically different languages.

To sum up, variability in the use of the verbal inflection is a robustly-observed behavior in the speech of L2ers as the rates of dropping the verbal inflection indicate. This variability appears as a developmental phenomenon in the speech of some learners as well as a persistent problem in the speech of others. The L1 does not seem to have an effect on whether L2ers experience variability or not but it might be a factor that determines its persistence. More characteristics of this phenomenon will be uncovered as we proceed in the following section.

As far as accounting for the phenomenon of morphological variability is concerned, to explain the observed L2ers' performance, researchers are faced mainly with two tasks as follows:

I. Locating the source of morphological variability that all L2ers go through.
II. Locating the source that leads to the persistence of morphological variability in some learners rather than others.

Indeed, many hypotheses have been advanced in the field of SLA addressing either of these tasks. The following section 2.3.2 deals with how L2 researchers view the relationship between surface morphology and the corresponding syntax in SLA and section 2.3.3 presents some of the main proposals on morphological variability along with illustrative studies to clarify the evidence that these accounts depend on to explain this phenomenon.

2.3.2 The Morphology and Syntax Relationship in SLA Research

The widely observed variable use of inflectional morphology has led to debates on the nature of L2 speakers' mental representations. As stated by Hawkins (2009: 232), "[m]uch of the debate about the nature of L2 speakers' knowledge of morphosyntax results from different interpretations that researchers have given to the relation between performance and underlying representation". In other words, there is disagreement among L2 researchers over the relationship that exists between surface morphology and its abstract syntactic representations. Does the use of surface morphology directly reflect the nature of the speaker's mental representations? Or, is it the case that L2ers unconsciously know more than what their use of the morphology shows? Positive answers to both questions come from two contrasting positions adopted by SLA researchers on the morphology-syntax relationship. The two positions are presented hereafter.

The first position on the morphology-syntax relationship is the 'morphology-before-syntax' approach (as termed by White, 2003b, p.184). In this position, it is believed that the functional categories are initially absent from the interlanguage grammars of L2ers and it is overt morphology in the linguistic input that drives their projection. This view is adopted by researchers (e.g., Vainikka and Young-Scholten, 1994, 1996, 1998, 2013; and Hawkins, 2001) who believe that L2ers come to the learning task with only some of the categories that comprise mature competence and other categories are built up subsequently gradually. Specifically, it is lexical categories (e.g., VP, NP, and AP) that are believed to be the starting point, with the functional ones (e.g., TP and CP) triggered by the input. In such a view, the absence or inconsistent use of a given functional morpheme in the speech of L2ers is taken to signal an absence of the corresponding grammatical category. With more exposure to the target language, learners notice that certain morphological forms (e.g., V+ed) occur in specific contexts and start using them, which triggers the instantiation of their corresponding syntactic categories. This is to say that functional morphology emerges in the speech of learners before the corresponding syntactic representations exist and the occurrence of overt morphology is what drives the emergence of syntactic categories. We will come back to the empirical evidence provided in support of this position when we present the Organic Grammar Hypothesis in 2.3.3.1 and the Modulated Structure Building Hypothesis in 2.3.3.2 below.

The second position on the morphology-syntax relationship the ‘syntax-before-morphology’ approach (termed so in White, 2003, p.182). It is believed here that the full range of lexical and functional categories exist in the interlanguage grammars of L2ers from the very beginning of SLA and the task of these learners is mapping these syntactic categories onto their overt manifestations (or vice versa); hence, it is the presence of these representations in the underlying grammars that drives the acquisition of the overt morphological forms. This position is adopted by SLA researchers who are proponents of the full access to UG view including those who assume a full L1 transfer (Full Access/ Full Transfer hypothesis e.g., Schwartz and Sprouse, 1996; Grondin and White, 1996) or no L1 transfer at all (Full access/No Transfer hypothesis; e.g., Epstein, Flynn & Martohardjono, 1996, 1998). Arguing against the view that maintains that the absence of the inflection from the speech of L2ers is an indicator of the absence of the corresponding grammatical category, Schwartz (1998: 44) asserts that “it seems equally plausible to say that the functional categories are indeed there – and even fully specified as in the L1 grammar – but just not overtly filled … initially”. Therefore, under this view, absence or inconsistent use of functional morphology is seen as a result of grammatical categories being not mapped into their manifestations (or vice versa) yet. We shall return to the supporting evidence for this view when presenting the Missing Surface Inflection Hypothesis in 2.3.3.4.

Therefore, the first position on the morphology-syntax relationship holds that the use of inflectional morphology directly reflects the nature of the underlying syntactic representations. In this case, the inconsistent use is believed to be a result of non-target-like representations. Three of the accounts of morphological variability presented in the following section maintain this position. These are the Organic Grammar Hypothesis (2.3.3.1), Modulated Structure Building Hypothesis (2.3.3.2) and Representational Deficit Hypothesis (2.3.3.3). The second position is that there is a dissociation between L2ers’ underlying syntactic knowledge and their use of inflectional morphology and therefore it assumes that the inconsistent use is not caused by non-target-like representations. This position is held in the other three accounts of morphological variability presented in the following section. These are the Missing Surface Inflection Hypothesis (2.3.3.4), the Prosodic Transfer Hypothesis (2.3.3.5) and the Phonological Reduction Hypothesis (2.3.3.6).

2.3.3 Theoretical Accounts of Morphological Variability

Various interpretations have been given to morphological variability experienced by L2ers. These can be divided into two groups according to whether they attribute the source of variability to syntactic or non-syntactic deficiencies. The syntactic accounts attribute the source of the problem to a temporary or permanent absence of the syntactic representations associated with the functional morphology. The Organic Grammar Hypothesis (2.3.3.1), the Modulated structure Building Hypothesis (2.3.3.2) and the Representational Deficit Hypothesis (2.3.3.3) fall in this group. On the other hand, the non-syntactic accounts maintain that the L2ers' use of surface forms under-represents their grammatical competence. Under this view, the missing inflection does not entail an absent syntactic category or feature; rather, the abstract syntactic categories are intact and fully present from the early stages of SLA and the source of the problem is caused by a processing failure as held by the Missing Surface Inflection Hypothesis (presented in 2.3.3.4) or a phonological deficiency as maintained by the Prosodic Transfer Hypothesis and the Phonological Reduction Hypothesis (presented in 2.3.3.5 and 2.3.3.6 respectively).

Presenting these hypotheses below, although their full proposals will be laid out, focus will be put on their account of morphological variability as well as their views on how linguistic development occurs for L2ers to overcome this target-deviant performance and what role the L1 plays in this regard.

2.3.3.1 The Minimal Trees Hypothesis/ Organic Grammar

Vainikka and Young-Scholten (1994, 1996a, b, 1998) propose the Minimal Trees Hypothesis, which has been developed into Organic Grammar in Vainikka and Young-Scholten (2005, 2007, 2011, 2013), in an attempt to describe the linguistic knowledge L2 learners initially have at the commencement of the learning task, hypothesising on the source of this knowledge and how it develops subsequently with more exposure to the target language. To show how this model accounts for variability in the use of inflectional morphology, presenting the full proposal is necessary. The more recent term, that is, Organic Grammar (OG), will be used to refer to this model.

OG is a structure building account of SLA. It proposes that L2ers go through a number of syntactic stages during which grammatical categories build up gradually. The initial state of SLA is held to be a VP stage, which comprises only lexical projections and lacks the functional ones. The source of this knowledge (lexical projections) is assumed to be the learners' L1s, from which only the lexical categories transfer but functional categories do not. This shows up in the early speech of learners as an absence of the phonological exponents of functional categories and the learners' utterances bearing properties of their L1, specifically the phrase headedness, which gives rise to the L1 word order. In the course of development, but within the same stage, the headedness of the phrase swiftly changes to match the input and, hence, the L2 word order appears.

In the next stages, functional projections (e.g., TP & CP), triggered by L2 input and its interaction with fully accessible UG, emerge in a gradual bottom-up manner. It is believed that the emergence of a given functional category is a consequence of the learners' identifying in the input elements associated with that category (morphology-before-syntax approach, see section 2.3.2 above). In Vainikka and Young-Scholten (1998), it is hypothesised that while the copula and modals are the triggers for a TP projection, the complementisers are the triggers for the CP category.

Under this conception of how categories are added to the interlanguage grammar, it is expected that a given category would not be absent and then become present at once; rather, it is more plausible to expect it to get instantiated in stages depending on the learners' ability to identify all of its phonological exponents in the input. For example, verb raising over adverbs and negation, modals, auxiliaries, tense and agreement markers on verbs are all properties believed to be reflections of features of the TP projection; the learners' identification of one or two of these properties might lead to the instantiation of the category but it does not necessarily follow that the category is fully specified for all features. Indeed, Vainikka and Young-Scholten (1994; 1996b) argue that the VP stage is followed by a stage that has a projection of finite phrase (FP) not TP. The FP according to the authors is an incomplete TP. Specifically, it is a TP underspecified for number, person or tense. This shows up in the L2 speech as frequent use of some exponent of the TP (e.g., modals and auxiliaries) and optional use of others (e.g., verb raising and inflectional morphology). Following the FP stage, a fully specified TP projection replaces it. Here, the manifestations of the category have become consistent and frequent in the L2ers' speech. However, in a recent work, Vainikka and Young-Scholten (2013) abandon the FP layer and this is no more held to be an intermediary stage between the VP and TP stages. Finally, the CP category is projected and gets specified for the appropriate features and their values.

Furthermore, this hypothesis holds that while the learner's L1 is partially involved in L2 acquisition, UG is fully available to assist L2ers in their task. As mentioned above the involvement of the L1 is restricted to a transfer of L1 lexical categories and their properties, which work as the starting point for L2ers. As more linguistic exposure takes place, the properties of the target language lexical categories start to replace the transferred ones. No L1 transfer is believed to occur beyond the VP stage. Subsequently, since UG is fully available, the building up of functional categories occurs based on the linguistic input from the target language and its interaction with UG. Therefore, although the instantiation of functional projections depends on the learner's encountering their phonological manifestations in the input, the sequence in which these projections are acquired is innately determined by UG.

To sum up so far, OG maintains that SLA goes into a number of developmental grammatical stages starting off with a lexical initial state and then functional categories are added incrementally. While lexical categories transfer from the learner's L1, functional categories do not. The triggers for projecting a functional category are its phonological manifestations in the input. Moreover, these categories do not project fully specified for all features at once but they go into stages where they can be specified for some features but not others until they become finally fully specified.

As for the phenomenon of morphological variability, the interpretation given by OG is the temporary absence of the functional categories and/or their features. According to OG, verbal morphology is expected to be absent from or inconsistent in the speech of L2ers at two stages. The first stage is the initial lexical state (i.e., VP stage). Here, the absence of the inflection from the speech reflects the absence of the associated functional category. The second stage where the inflection is expected to be produced inconsistently is a stage where the associated category is not fully specified for its component features (i.e., FP stage). Inconsistency in inflection suppliance here is due the corresponding projection being underspecified for tense, person and number features. However, at both stages the inflection might occur sometimes on verbs; such occurrences are interpreted by Vainikka and Young-Scholten as inflections being part of the lexical items as unanalysed forms. The verbal morphology will start to be used consistently only when the corresponding underlying functional category is fully specified for its features.

Evidence in support of the OG comes from the production data collected longitudinally and cross-sectionally from 28 naturalistic adult L2ers of German, speakers of different L1s, i.e., Turkish, Korean, Spanish and Italian reported in the early work of Vainikka and Young-Scholten (1994, 1996b) (see also Vainikka and Young-Scholten, 2011 and 2013, which provide evidence for OG stages from three adult English-speaking naturalistic L2ers of German). The data support the developmental stages posited in the OG. What is of interest to us here is the use of TP exponents by learners who are placed at the VP or FP stages.

Starting with the VP stage, eleven of Vainikka and Young-Scholten's (1994, 1996b) informants were identified as being at this stage. The speech of these informants as summarised in Vainikka and Young-Scholten (1996a: 17-18) showed that (1) verbs were rarely raised over adverbs and negation (less than 15%) where raising is obligatory in target grammar, (2) modals and auxiliaries were produced by some of the informants but infrequently (none produced more than 8%) and (3) the use of the non-inflected verb was around 61% of the time. Given that the properties of verb raising, auxiliaries, modals and inflected verbs are associated with a TP projection, the authors interpreted the non-target like performance of their informants as an absence of such functional category. Therefore, these findings together, according to the authors, are indicative of the absence of the corresponding syntactic category.

The next stage is an FP stage. Nine of Vainikka and Young-Scholten's informants were placed at this stage. The speech of those learners showed the following: (a) verbs were raised over adverbs and negation "about half of the time" (Vainikka and Young-Scholten, 1998: 25), (b) the use of modals and auxiliaries became more frequent and (c) the use of the non-inflected verb was around 57% of the time.[6] Comparing the performance of these learners to those placed at the VP stage, it can be observed that although verb raising, modals and auxiliaries become more productive, the use of the verbal inflection remains the same. The authors argue that although this performance indicates that a TP is projected, this category does not seem to be fully specified as the

[6] Findings in b and c are based on Vainikka and Young-Scholten (1996a, table 4:.22).

verb raising is still optional and the use of agreement and tense marking is inconsistent. Therefore, this suggests the presence of an incomplete TP projection, thus, FP.

Vainikka and Young-Scholten (1994, 1996b) also provide evidence for L1 transfer during the VP stage and for its lack during subsequent (functional) stages. The authors reported that their informants at the VP stage produced constructions in L2 German that bore the headedness characteristic of their native languages, but those who were at the FP stage did not do so. Vainikka and Young-Scholten explain that Turkish and Korean, similar to German, have a head-final VP, but Italian and Spanish, unlike German, have a head-initial VP. The data from their informants at the VP stage showed that while the Turkish and Korean informants produced head-final VPs in 95% of the time, their Spanish and Italian counterparts predominantly produced head-initial VPs. This was interpreted as evidence for L1-transfer of the VP projection properties. However, the case was different at subsequent stages as similar transfer at the level of functional categories was not attested. Although functional categories in Spanish and Italian, similar to German, are head-initial and in Turkish and Spanish, different from German, are head-final, the authors observed that all of their (different L1) informants produced head-initial functional projections.

All in all, the hypothesis advanced by Vainikka and Young-Scholten attributes the omission or inconsistent use of surface morphology to the temporary absence of corresponding functional categories and/or their features. Hence, this morphological variability is looked at as a phenomenon experienced by learners whose interlanguage is in the course of development and has not projected a TP category. As L1 transfer is believed to occur only initially at the VP stage, no L1 effect is thought to be involved in this phenomenon. Moreover, since UG is fully available in SLA, L2ers are expected to converge on the target grammar and thus overcome this target-deviant performance.

2.3.3.2 The Modulated Structure Building Hypothesis

Another hypothesis that addresses the phenomenon of morphological variability is the Modulated Structure Building Hypothesis (MSBH) advanced by Hawkins (2001). In this hypothesis, Hawkins proposes a modulated version of the Minimal Trees Hypothesis (termed Organic Grammar (OG) above and, thus, below) of Vainikka and

Young-Scholten (1994, 1996b). Similar to OG, the MSBH maintains that L2ers' initial grammars consist of only lexical projections bearing properties of the learner's L1 and functional projections are added subsequently and gradually based on the positive evidence available in the input. Therefore, resembling OG, the MSBH locates the source of morphological variability at the temporary absence of the corresponding syntactic categories as both hypotheses share the same perspective on learners' syntactic development. The point of divergence between the two hypotheses however is that while OG assumes that the L1 effect is restricted to an initial transfer of the L1 lexical categories and their properties, the MSBH claims that the L1 effect is not restricted to this kind of transfer but also it becomes evident "once functional categories are established in the L2 grammar… and even then only at the relevant points of development" (Hawkins, 2001: 74). The relevant point of development according to Hawkins is when the learner's grammar requires a representation for a specific linguistic property for which there is positive evidence in the input. Here, if the learners' L1 system has the same property, it precipitates its acquisition in L2, giving advantage to such learners over other learners whose L1s lack the same property.

Hawkins (2001) provides supporting evidence for the MSBH from the findings of some studies conducted on the acquisition of L2 English. One such study cited by Hawkins (2001) is Stauble (1984), which tested the use of English copula *be*, auxiliary *be*, the progressive marker *–ing*, verbal agreement inflection *–s* and the past tense regular and irregular morphology by six Spanish and six Japanese L2ers of English at three different proficiency levels: low intermediate, intermediate and advanced (two informants from each language at each level). The results of Stauble's study showed that the low intermediate level learners used thematic verbs in the bare form (talk) and -*ing* form (talking) interchangeably and a high proportion (above 50 %) of these forms was used in a non-target-like manner. Moreover, in the performance of the same learners, the target-like use of auxiliary *be* was very low (below 20% for all informants) and the agreement and past tense marking on thematic forms hardly existed in their speech. Hawkins speculates that these findings are indicative of the absence of the corresponding syntactic categories. In contrast to the performance of the low level learners, Stauble's advanced level informants performed considerably better in the use of *–ing,* regular past tense and 3rd person agreement inflections showing also a noticeably lower proportion of non-target-like use. This performance was also coupled with a high rise in a target-like use of auxiliary *be*, indicating the emergence of the corresponding syntactic projection in the mental grammar of those learners. According to Hawkins, these observations are compatible with a structure building account (both OG and MSBH).

As for the 'Modulated' part of Hawkins' hypothesis, which constitutes the point of divergence from OG, Hawkins asserts that Stauble's study supports this proposal. The study findings revealed that the Spanish informants at the advanced level were more accurate than their Japanese counterparts in their use of 3rd person agreement inflection. Hawkins holds that this is suggestive of an L1 influence because while Spanish has a rich system of subject-verb agreement, Japanese does not, a factor that gives an advantage to the former group over the latter in their acquisition of this property in L2 English. Hawkins goes further to explain why this L1 advantage in this regard appeared in the performance of advanced level informants but not in the performance of the informants at the two lower levels. He proposes that the L1 influence does not occur until relevant points of development. As subject-verb agreement is in a specifier-head relation within the Inflection Phrase (IP: TP in current generative accounts), the relevant point of development is when the IP is fully established in the mental grammar. For the low intermediate and intermediate level informants, their use of the auxiliary *be* and past tense morphology did not indicate that they had acquired the category, rather they were in the process of acquiring it. In contrast, the results of the advanced level informants indicated that their mental grammars had projected this category and thus the time had come for establishing a specifier-head relation.

Overall, According to the MSBH, morphological variability is caused by the absence of underlying syntactic representations, which are built gradually based on the positive evidence in the input. The building up of functional categories is influenced by the learners' L1s as that if a certain property exists in the L1, its acquisition in the L2 becomes easier.

2.3.3.3 The Representational Deficit Hypothesis

L2 research has also shown that some L2ers persist in having problems with certain morphosyntactic properties at advanced proficiency levels (e.g., Hawkins and Chan, 1997; Hawkins and Liszka, 2003; Hawkins and Hattori, 2006) and even after long naturalistic exposure to the target language (e.g., Lardiere 1998a, 1998b; Franceschina, 2001; Tsimpli, 2003). One explanation that has been put forth to explain such persistent problems is that the relevant syntactic representations are impaired because specific linguistic features (i.e., uninterpretable features, see section 2.2.1) are subject to a critical period after which they become not acquirable. It is this proposal that is termed the Representational Deficit Hypothesis (RDH).

The RDH was advanced at first within the Principles and Parameters approach by Tsimpli and Roussou (1991), who suggest that where L1 and L2 differ in terms of parametric values associated with functional categories, difficulties arise and no re-setting to the L2 value is expected. Along the same lines, Hawkins and Chan (1997) propose the Failed Functional Features Hypothesis (FFFH). This version of the RDH attributes morphosyntactic problems to the absence of features related to functional categories from the interlanguage grammars. Hawkins and Chan argue that a critical period affects such features and therefore when not activated in early linguistic experience(s), they become inaccessible in post-puberty.

A more recent formulation of the RDH (e.g., Hawkins and Liszka, 2003; Tsimpli, 2003; Hawkins and Hattori, 2006; Tsimpli and Dimitrakopoulou, 2007), based on a Minimalist conception of the language faculty, proposes that uninterpretable features which are not available in the L1 or in any linguistic experience in childhood, become inaccessible to post-puberty L2ers. Tsimpli and Dimitrakopoulou (2007: 224) argue that "uninterpretable features are subject to critical period constraints and, as such, they are inaccessible to [older] L2 learners." Consequently, this deficit in the interlanguage grammars results in persistent problems with morphosyntactic properties such as variability in the production of functional morphology.

The evidence for the RDH comes from research examining the acquisition of uninterpretable features by adult L2ers who do not have them in their native language. Tsimpli (2003) is one exemplar study. Tsimpli examines the acquisition of determiners in definite and indefinite contexts, pronominal object clitics, as well as tense, agreement, mood and modality marking on verbs in L2 Greek. The informants of this study are six speakers of L1 Turkish and Russian learning the target language in a naturalistic setting. At the time of the study, they had been living in Greece for about nine years during which they had been in a Greek-speaking environment using the L2 in an everyday basis. Oral spontaneous production of each informant was recorded in a 45-minute interview.

Greek has definite and indefinite determiners which are inflected for case, number, and gender, and object clitics which are also inflected for case and *phi*-features (e.g. person, number and gender). Moreover, tense distinctions and agreement are morphologically marked on verbs, and Future and Mood are expressed on modal particles. Syntactically speaking, Tsimpli (2003) assumes that definite determiners are associated with uninterpretable features as they serve a purely grammatical function and are void of semantic content whereas indefinite determiners are linked with interpretable features in virtue of their "inherent feature-specification for referentiality" (p. 332). A similar assumption is held for object clitics; "specifically, uninterpretability of case and *phi*-features is associated with 3rd person object clitics whereas 1st and 2nd person clitics are distinct in that the person feature in their case is interpretable" (Tsimpli, 2003: 332). On the other hand, Turkish and Russian completely lack a determiner system distinguishing between definite and indefinite articles, and both languages allow for null objects[7]. In addition, Turkish and Russian do not have any pronominal object clitics. With regard to verbal morphology, the L1s of the informants and their L2 are similar, with only one difference between Turkish and the other two languages in that in Turkish mood and modality are expressed on suffixes on main verbs whereas these are encoded on modal particles in Greek and Russian[8].

Given the similarities and differences between the L1s and the L2 of the informants, the RDH predicts the following: (1) Definite determiners will be problematic in contrast to their indefinite counterparts due to the interpretability distinction and the unavailability of the uninterpretable features in the L1s of the informants, (2) pronominal 3rd person object clitics will be problematic in contrast to the 1st and 2nd person clitics due to the same reasons and (3) verbal morphology will not be problematic because their corresponding formal features are present in the L1 of the informants.

The results of the study showed that the three predictions were met. First, there was a statistical difference between the use of indefinite and definite articles for all six

[7] This is associated with the absence of the D system (see Tsimpli, 2003: 334)

[8] This difference between Turkish and the other two languages is not relevant to the interpretability of underlying features.

informants; the proportion of accuracy in indefinite articles was around 90%, reaching above 95% for four of the informants, whereas the accuracy rate for the use of definite articles was around 60%, with only two informants scoring around 70% and one informant under 15%. Second, there was a significant difference between the use of 1st and 2nd person object clitics in one hand and 3rd person object clitic on the other for all informants; accuracy on 1st and 2nd person clitics was around 90%, reaching above 95% for three informants, and, in contrast, for the 3rd person clitics, none of the informants scored above 73%, with three scoring under 40%. Third, all six informants performed in a native like manner (between 99% and 100%) in Tense, Agreement, Mood and Modality marking. According to the author, these results are compatible with the RDH as problems arose only in properties associated with uninterpretable features not instantiated in the informants' L1s.

In summary, the RDH proposes that missing syntactic features from the interlanguage grammar of L2ers cause persistent inconsistency in the use of functional morphology. The absence of specific features from the interlanguage grammars is due to the inactivation of these features from the universal inventory of features in the language experience(s) before the critical period.

The hypotheses presented so far in this section locate the source of morphological variability in the underlying syntactic representations. While OG and MSBH address the variability experienced by all L2ers in the process of development, the RDH claims to account for the residual variability. However, some L2 researchers, (e.g., Lardiere, 1998a, b; Prévost and White 2000a, b; Prévost 2003) reject the syntactic deficiency account of the phenomenon and propose alternatively a number of non-syntactic accounts. These are presented in the rest of this section.

2.3.3.4 The Missing Surface inflection Hypothesis

Contra the structure building accounts presented above (OG & MSBH), some researchers (e.g., Schwartz & Sprouse, 1994; 1996) maintain that the L2ers start their learning task with the full set of syntactic categories already in place, which are thought to be transferred from the L1 linguistic system. Moreover, they have full access to UG, which enables them to restructure their grammars to match the target language grammars based on the received input (Full Transfer/Full Access hypothesis). Under this proposal, the absence (or variable use) of verbal morphology from the L2 speech does not indicate absence of the corresponding underlying syntactic representations; it is rather believed to be, as Haznedar and Schwartz (1997: 266) put it, "a problem with just **realizing** the morphological form of finite verbs".

Haznedar and Schwartz (1997) reached this conclusion based on examining data collected from a Turkish-speaking child early learner of L2 English. They found that despite the inconsistent use of verbal morphology in their informant's speech, other related syntactic properties such as overt subjects and nominative case marking were accurately supplied. Building on Chomsky's (1995) proposal that overt subjects and nominative case marking are licensed by the tense head (T), Haznedar and Schwartz considered their informant's performance an indicator of the presence of the TP projection.

Similarly, Lardiere (1998a, 1998b, 2000) argues that the use of surface morphology under-represents L2 speakers' underlying knowledge and, therefore, the inconsistent use of morphological items should not be taken as a reflection of absent or impaired representations; rather, according to Lardiere, the issue is that "for L2 acquirers, the problem lies in figuring out how (and whether) to spell out morphologically the categories they already represent syntactically, i.e., the 'mapping problem'" (Lardiere, 2000: 121).[9]

Building on the proposals advanced by Haznedar and Schwartz (1997) and Lardiere (1998a, b, 2000), Prévost and White (2000b) proposed the Missing Surface Inflection Hypothesis (MSIH). The MSIH maintains that L2 learners do have unconscious knowledge of the functional projections including tense and agreement underlying representations; however, the problem lies in the difficulty that some L2 learners

[9] More recently Lardiere (2008, 2009) proposed the Feature Reassembly Hypothesis. In this hypothesis, Lardiere suggests that since L2ers, different from L1ers, come to the task of learning with a set of fully assembled grammatical categories and which might differ in how they are assembled in L1 and L2, "this will require that the learner reconfigure or remap features from the way these are represented in the L1 into new formal configurations on possibly quite different types of lexical items in the L2" (Lardiere, 2009: 175). Under this proposal, variability (as well as other target-deviant phenomena) experienced by L2ers is thought to be a result of the process of reassembly that learners go through during acquisition of the target language.

experience in the realization of surface morphology. Adopting the Distributed Morphology (DM) framework (Halle and Marantz, 1993), Prévost and White clarify the mechanism through which non-finite verbs appear in finite contexts in the speech of L2ers. In the DM framework, it is assumed that lexical items are inserted into syntactic nodes according to a competition process that allows the item with the most matching features to those held by the syntactic node to be inserted. Infinitival verbs are default forms underspecified for finiteness in contrast to other verb forms, which are specified for finiteness. When a learner, for example, needs to produce a finite verb, it is supposed that the features held by the terminal node in the syntactic structure are fully specified and therefore, the lexical form with most matching specifications is expected to be retrieved to access the computation for feature checking. However, according to Prévost and White (2000b: 129), "due to processing reasons or to communication pressure", a failure in the retrieval of the appropriate form might occur and as a result a less specified form is inserted instead.

Prévost and White's (2000) study provides evidence for the MSIH. In this study, the authors investigated spontaneous oral production data obtained from two adult learners of L2 French (L1 Arabic, Abdelmalek and Zahra) and two adult learners of L2 German (one L1 Spanish, Ana, and one L1 Portuguese, Zita). The data from the two L2 German learners cover a period of two years starting three months after their arrival in Germany and the data from the two L2 French learners cover a period of three years starting one year after their arrival in France. Examining these learners' speech, Prévost and White found the following: (1) Non-finite verbs were found in place of finite verbs but the opposite rarely occurred, and (2) when agreement on verbs was supplied, it was appropriate to a high extent. These findings are based on the results summarised in Table 2.1.

Table 2.1: Distribution of finite and non-finite verb forms and the appropriateness of finite verb forms to subjects (thematic and non-thematic verbs) (Based on Prévost and White, 2000, Tables 4, 6, 8 and 9: 114-124)

	Non-Finite V in Finite Contexts	Finite V in Non-Finite Contexts	Appropriateness of Finite Verb Forms to subjects
Zahra (L2 French)	217/837 (25.9%)	2/151 (1.3%)	708/749 (94.5%)
Abdelmalek (L2 French)	237/914 (25.9%)	14/234 (5.6%)	711/742 (95.8%)
Ana (L2 German)	36/343 (10.5%)	5/64 (7.2%)	561/646 (88.1%)
Zita (L2 German)	72/332 (21.7%)	2/70 (2.8%)	368/419 (87.8%)

These findings led Prévost and White to conclude that the TP projection is represented in these learners' underlying grammars because it is the presence of [finite] feature that leads these learners to differentiate between finite and non-finite verb forms and [Agr: person] and [Agr: number] features that enable them to supply the correct verb forms with the appropriate subjects. These features are assumed to be hosted in the TP. Therefore, this is held as evidence for the authors' proposal that variability is caused by a missing surface inflection.

Support for the MSIH comes from other studies showing that variability occurs in the speech of learners at the same time while evidence for the presence of TP projection exists. Prévost (2003) provides such evidence. In this study, Prévost examined spontaneous and elicited production data collected from four English-speaking adult L2 learners of French. The informants scored at the beginner level in a placement test prior to their admission to a language program at Laval University in Canada. After one or two months of the start of their program, they were interviewed on a monthly basis for a period ranging between two and seven months. Prévost found that all learners alternated between using the finite and non-finite verb forms in finite contexts: the percentages of non-finite verbs varied across learners, ranging between 3% (34/1058) and 30% (61/152).[10] Despite this variability, however, evidence showing that the TP category is projected exists in these learners' speech. The author observed that all informants frequently used nominative-case-marked subjects even in utterances containing non-finite verb forms. Based on Chomsky's (1995) proposal that overt subjects and nominative case marking are licensed by Tense, the TP projection does exist in these learners' underlying representations.

Likewise, Herschensohn (2001), investigating the L2 French of two English-speaking adolescents who were at an intermediate level of proficiency, also found similar performance. Herschensohn's informants alternated between using finite and non-finite verb forms in finite contexts. It was observed that the non-finite verbs occurred in clauses filled with nominative-case-marked clitics and DP subjects. Again, this suggests that the [Case] feature, hosted in T, is represented in these learners underlying grammars.

[10] This is based on Table 3 in Prévost (2003:.374)

In summary, the MSIH assumes that morphological variability is caused by a processing failure between overt morphology and their syntactic representations. Syntactic categories are fully represented in the interlanguage of L2ers from the very beginning and their task is restricted to the learning of surface forms. Variability is therefore seen as a phenomenon experienced by learners in the course of acquisition. Finally, as UG is fully accessible in SLA according to this hypothesis, learners are expected to converge on the L2 grammars and the presence or absence of certain properties in the L1 makes no difference in SLA.

Although the MSIH provides an account for the morphological variability experienced by all L2 learners in the course of linguistic development, it cannot explain why this phenomenon persists into, at least, advanced stages of proficiency in some learners rather than others. This has led proponents of the MSIH to advance additional proposals to account for this performance. These proposals are mainly phonological as we shall see below.

2.3.3.5 The Prosodic Transfer Hypothesis

To account for the residual difficulty some learners, particularly L1 speakers of Chinese, rather than others have in the use of inflectional morphology of L2 English, Goad, White and Steele (2003) propose the Prosodic Transfer Hypothesis (PTH). The PTH states that prosodic constraints transferred from the learners' L1 are responsible for the failure of L2ers of English to supply inflectional morphology in a consistent manner. It is held that languages differ as to how they prosodify functional elements and, therefore, transfer of prosodic structures causes problems with producing the L2 functional elements, specifically when there is a mismatch between the L1 and L2 of learners in this domain. Goad et al.'s proposal is built upon assumptions on (1) the structure of prosodic constituents (2) constraints dominating them and (3) differences between the learners' L1 and L2 in this regard. I will deal with these in turn hereafter.

A key idea from Prosodic Phonology that Goad et al.'s proposal adopts is the Strict Layer Hypothesis (SLH; e.g., Nespor and Vogel 1986; Selkirk 1986). The SLH suggests that the prosodic structure in languages is controlled by a strict layering constraint, which ensures that prosodic constituents are structured hierarchically with each level of the structure directly dominated by the immediate higher level. The hierarchical structure presented in Figure 2.2 is widely adopted in Prosodic Phonology.

Figure 2.2: the (partial) prosodic hierarchy adopted in Goad et al. (2003)

Phonological Phrase
|
PWd (Prosodic word)
|
Ft (Foot)
|
σ (Syllable)

Goad et al. explain that the prosodic structure in languages was initially thought to respect the strict layering constraint determining that the prosodic constituents are structured in the order posited in Figure 2.2. For a long time, this constraint was held to be inviolable. However, Selkirk (1997) proposes that the strict layering should be understood in terms of four violable constraints, in line with the Optimality Theory approach (Prince and Smolensky, 1993). Two of the constraints proposed by Selkirk (1997) are relevant to Goad et al.'s proposal. These are as follows (from Goad et al., 2003: 247).

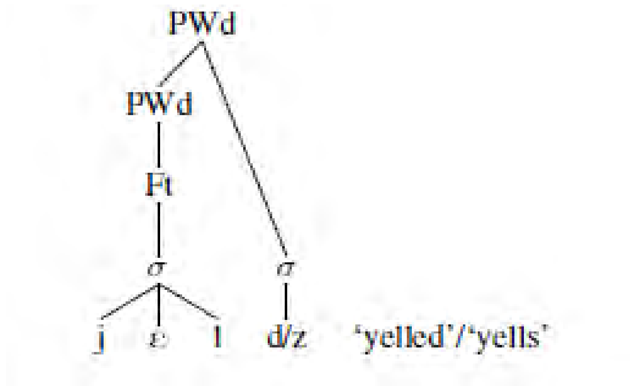
1) EXHAUSTIVITY (EXHAUST): No C^i immediately dominates a constituent C^j, $j< i-1$ (e.g., no PWd immediately dominates a σ)

2) NONRECURSIVITY (NONREC): No C^i dominates a C^j, $j = i$ (e.g., no PWd dominates a PWd)

The idea that these constraints are violable in languages is essential to the PTH as we will see below.

Goad et al. argue that, in English, suffixes such as the past tense marker *-ed* and the 3^{rd} person agreement marker *-s* are adjoined directly to the external PWd simultaneously violating the EXHAUST and NONREC constraints. This is schematically represented in Figure 2.3.

Figure 2.3: Prosodification of past and agreement inflection (adopted from Goad et al., 2003: 248)



However, the case is different with irregular pseudo-inflected past forms (e.g., wept). In such forms, Goad et al. argue that past tense marking is prosodified PWd internally violating only the EXHAUST constraint. The same is true for uninflected monomorphemic words that have word-final consonant clusters and similar in shape to regularly inflected verbs (e.g., adapt).

In sum, it is argued that regular past and verbal agreement are prosodified PWd externally rendering an adjunction structure that violates the EXHAUST and NONREC constraints. By contrast, only the former constraint is violable in prosodic structures of irregular past tense forms and monomorphemic words ending with consonant clusters.

Given these assumptions about the prosodic structure of English inflected verbs, Goad et al. suggest that L2ers of English whose L1s do not allow prosodic structures violating these constraints in a similar fashion would encounter difficulty in supplying the L2 inflection until they acquire the prosodic adjunction structure.

In this proposal, if the native language of L2ers of English prosodifies inflectional morphology in the English-like fashion, no problems are expected to arise in supplying past and agreement morphology. However, in case of L1 difference in morphology prosodification, Goad et al. maintain that two patterns of behavior are expected. The first expected behavior is that some learners might comprehensively delete English inflections in oral production due to their realization of the mismatch between what their grammar permits and the external-word adjunction analysis of English inflection. The second expected behavior is that other learners might variably use the L2 morphology. This occurs when other prosodic structures in the L1 can be used to accommodate the structure required for the English inflection.

Goad and White (2006) propose that although differences between the L1 and L2 of speakers cause difficulties in the production of the L2 inflection, this can be overcome ultimately. That is, the L2 prosodic structure can be acquired and hence this will lead to consistent production of the L2 inflection.

All in all, the PTH proposes that L1 prosodic constraints transfer to the interlanguage grammar and control L2er's production. Accordingly, prosodic structures not present in the L1 are a source of difficulty. This difficulty, however, can be overcome ultimately when L2ers acquire the L2 prosodic structure.

The empirical evidence for PTH will be reviewed in Chapter three (section 3.2.1). Another phonological account of morphological variability is presented in the following section.

2.3.3.6 The Phonological Reduction Hypothesis

Lardiere (1998a and b; 2003), as demonstrated above (section 2.3.3.4), maintains that morphological variability is caused by a mapping problem between overt morphology and the corresponding syntactic representations. Lardiere also claims that the omission of the inflection is aggravated by a phonological deficiency. Building on Bayley's (1991, 1996) observation that in past tense verb forms his Chinese informants tended to reduce final consonant clusters (by *-t/d* deletion), Lardiere (1998a, 2003) proposes the Phonological Reduction Hypothesis (PhRH), which holds that morphological variability is promoted by a consonant cluster reduction. This reduction is believed to be due to constraints imposed by the learner's L1. Lardiere (1998a: 20-21) puts it as follows:

> We can further imagine that an essentially morphophonological mapping procedure would be especially vulnerable to 'derailment' from a variety of post-syntactic or extra-syntactic factors, such as phonological transfer from the L1. For example, neither Hokkien nor Mandarin has final consonant clusters – a fact undoubtedly relevant to the likelihood of an adult native Chinese speaker producing (or perhaps even perceiving) regular, unstressed past tense suffixes on English verbs ending in consonants.

It is clear that Lardiere here contends that since Chinese disallows consonant clusters, a transfer of this constraint would prevent Chinese L2ers of English from consistently producing inflections which create consonant clusters word finally when added to verb stems (e.g., learned /nd/; talked /kt/). We will return to Lardiere's evidence in support of the PhRH in Chapter three (section 3.2.1)

The following is a summary of the six accounts of morphological variability presented in this section:

- **Organic Grammar (OG)**: Morphological variability is viewed here as a phenomenon experienced by L2ers in the course of acquisition. Its source is held to be a temporary absence of the corresponding syntactic categories, which are built gradually with more exposure to the target language. The L2ers' native language does not play any role in this regard.

- **The Modulated Structure Building Hypothesis (MSBH)**: Similar to OG, this hypothesis views morphological variability as a developmental phenomenon and it is caused by a temporary absence of the corresponding syntactic representations. However, different from OG, this account holds that the presence of certain properties in the L1 speeds up their acquisition in the L2 at relevant points of development.

- **The Representational Deficit Hypothesis (RDH)**: This hypothesis accounts for the persistent aspect of morphological variability. It gives the native language the prominent role in explaining the phenomenon in SLA; it maintains that uninterpretable features which do not exist in the L1 or any linguistic experience in childhood become permanently unavailable for adult L2ers, so they cause persistent problems. .

- **The Missing Surface Inflection Hypothesis (MSIH)**: Contra OG, MSBH and RDH, which attribute morphological variability to a syntactic deficiency, this hypothesis proposes that the syntactic structure is fully present from early on in SLA (the source is the L1 or UG) and the phenomenon is caused by a processing problem in the production of inflectional morphology. This processing problem in turn is thought to be caused by communicative pressure. Furthermore, UG is believed to be fully available in SLA and, therefore, whether a specific property is present or absent from the L1 does not make a difference in SLA.

- **Prosodic Transfer Hypothesis (PTH)**: This hypothesis attributes the source of morphological variability to phonological constraints transferred from the L1. It is claimed that English past tense and verbal agreement inflections attach externally to the prosodic word creating an adjunction structure; if the L2ers' native language does not allow such a structure, difficulties in the production of the inflection arise. However, these difficulties can be overcome when the L2 structure is acquired.

- **The Phonological Reduction Hypothesis (PhRH)**: This hypothesis holds that L1-transferred phonological constraints at the syllable structure promote L2er's difficulty in the production of the inflection. English past tense and verbal agreement inflections sometimes create a consonant cluster in a word final position (i.e., travelled /ld/; travels /ls/); such consonant clusters are thought to cause difficulties to L2ers whose L1s disallows them, which manifests in variable production of the inflection.

Finding out the source of morphological variability experienced by adult L2ers is the primary objective in this thesis. As mentioned before, this phenomenon has been subject to scrutiny benefiting primarily from production and, to far less extent, perception data. A growing body of research has recently started investigating how L2ers process functional morphemes in real time comprehension, but little use has been made of its findings to understand the reason of this target-deviant phenomenon. In this thesis, I take the stance that the source of morphological variability and the mental representations underlying such phenomenon can be better understood through inspecting how L2ers process the relevant morphological elements. Therefore, the remainder of this chapter focuses on L2 processing of morphosyntax.

2.4 Second Language Processing of Morphosyntax

As will be detailed in Chapters three and seven, in addition to the production and perception of English past tense and verbal agreement morphology, this thesis also explores the processing (during perception) of the same properties in SLA. This section therefore lays the foundation for this exploration. It starts with a discussion of how processing data can be used to provide evidence for or against the hypotheses presented in section 2.3. It will then briefly introduce the field of L2 processing, focusing on adult

L2 processing of morphosyntax and how it is different from native processing. After that, two factors that are relevant to this thesis and thought to generate differences in processing between non-native and native speakers, namely L2 proficiency and L1 transfer, are discussed before the chapter comes to a close.

As mentioned before in this chapter, the focus of the generative approach to SLA has been on explaining the nature of the linguistic competence of L2ers and how it develops over time. In another related field, namely psycholinguistics, research has focused on characterising the processing mechanisms involved in language production and comprehension and on factors that might affect those mechanisms. Although the two approaches differ in their focus, they have an overlapping relationship; on the one hand, successful processing takes place only if the relevant grammatical knowledge exists, and, on the other hand, for successful acquisition of grammatical knowledge, the appropriate parsing mechanisms must be available (Clahsen and Felser, 2006). It is self-evident that investigating how L2ers process language informs us about their grammatical knowledge and provides experimental evidence in support of or against theoretical proposals on its nature. Clahsen (2007: 97) asserts that "psycholinguistic findings may favour one theory of grammar over its alternatives and thus help to resolve theoretical linguistic controversies". The theoretical controversies this thesis seeks to resolve are those on the nature of grammatical knowledge driving the inconsistent use of inflectional morphology, as discussed above.

In the previous section, it was shown that as far as morphological variability is concerned, the controversy is over whether the syntactic representations underlying this phenomenon are absent or, rather, present but variability is caused by other non-syntactic factors. This is the particular controversy that processing data can resolve. Since successful processing of a specific morphosyntactic property cannot take place without the presence of the underlying syntactic representations, if successful processing (during perception) is attested at the same time as the production is variable, it would be safe to conclude that morphological variability is not caused by a syntactic deficiency. On the other hand, if successful processing (during perception) is not attested at the same time as production is variable, the source of morphological variability could be syntactically driven. Nevertheless, it should be noted that the latter case is not as straightforward as it appears to be. This is because unsuccessful processing might occur not only because of syntactic deficiencies but also some other factors might have such effect. This section (in 2.4.1) shall discuss two such factors in the aim of a better understanding of processing data.

Processing research has relied on data from a variety of time-sensitive or online psycholinguistic methods such the measurement of latencies in reading, or in response to, linguistic stimuli or physiological measures of event-related brain potentials (ERPs) or eye movements during real-time comprehension. Such data are particularly interesting to linguistic research because they show how grammatical representations are formed in real time of speech production and comprehension and “reduce the possibility of participants relying on their explicit or metalinguistic knowledge, compared to the more commonly used offline tasks” (Clahsen, Felser, Neubauer, Sato and Silva, 2010: 22).

As in SLA research, a common practice in L2 processing research is to measure L2ers’ performance in a specific linguistic experiment against data collected from native speakers of the target language. This is to examine to what extent the L2ers’ performance approaches that of the native speakers, which is taken as indication of success. In the previous section, we showed how adult L2ers experience difficulties with the production of morphosyntactic properties, particularly inflectional morphology, diverging from the native speakers’ patterns. Likewise, L2 processing research has recently revealed that such difficulties in morphosyntax do extend to processing during listening or reading showing differences between this population and their native counterparts (Clahsen et al., 2010). These differences appear as adult L2 learners, compared to native speakers, being less sensitive to morphosyntactic information during speech processing. The following is an exemplary study of such differences.

In an eye-tracking study, Keating (2009) examined the processing of Spanish gender agreement by 12 beginning, 14 intermediate and 18 advanced English-Spanish late learners and 18 Spanish native speakers. Keating found differences between the three non-native groups and the native one showing up as insensitivity to gender agreement violations by the non-native groups. The author gave his informants a set of sentences that involved gender agreement between the noun and the adjective. The distance between the noun and its adjective was controlled for as the test sentences included sentences of three types manipulating the position of the adjective: (1) the adjective occurs with the noun within the determiner phrase as in (2.5a), (2) the adjective occurs outside the determiner phrase and within the verb phrase of the matrix clause as in (2.5b) and (3) the adjective occurs outside the determiner phrase and within the verb phrase of the subordinate clause as in (2.5c).

(2.5)

a. [IP Una casa pequeña [VP cuesta mucho en San Francisco.]]
a house.**FEM**[11] small.**FEM** costs much in San Francisco
A small house costs a lot in San Francisco.

b. [IP La casa [VP es bastante pequeña y necesita muchas reparaciones.]]
the house.**FEM** is quite small.**FEM** and needs much repair
"The house is quite small and needs a lot of repairs."

c. [IP Una casa [VP cuesta menos [CP si [VP es pequeña y necesita reparaciones]]]].
a house.**FEM** costs less if it small.**FEM** and needs repair
"A house costs less if it is small and needs repairs."

(from Keating 2009: 505-506)

Grammatical (involving correct gender agreement as in 2.5) and ungrammatical (involving incorrect gender agreement) sentences were included in the test stimuli. Eye movements of the participants during the reading of the stimuli were recorded. Sensitivity to violations was measured by monitoring the length of fixation on ungrammatical adjectives compared to grammatical ones. The results showed that the native speakers were sensitive to the gender agreement violations as they fixated longer on ungrammatical adjectives as compared to grammatical adjectives in the three sentence types. In contrast, the beginning and intermediate non-native speakers were insensitive to the violations in the three types of sentences. The advanced learners performed similarly to native speakers in only the first type of sentences, that is, where the adjective occurred close to the noun within the determiner phrase. Therefore, although progress with rising proficiency was observed in the non-native groups' performance, they did not seem to have developed a native-like processing for all sentence types even at the advanced level.

That non-native speakers' processing of morphosyntactic properties is different from native speakers' processing is not debated, but the factors that cause such differences are. The remainder of this section will focus on two of such factors.

[11] FEM = feminine

2.4.1 Possible Factors Affecting L2 Processing

Attempting to explain the differences between non-native and native speakers' processing, researchers have identified a number of factors that they believe might generate differences between the two populations. Age of acquisition, L2 proficiency, L1 transfer, immersion experience and working memory are factors that are thought to possibly affect L2 processing (for a review see e.g., Dussias and Pinar, 2009). Two factors that are of particular interest to this thesis are L2 proficiency and L1 transfer and, therefore, we elaborate on them below.

2.4.1.1 L2 Proficiency

In the field of SLA, it has been found that L2ers with higher general proficiency, as measured by standardised tests or other means such as the length of exposure to the target language, perform better in linguistic experiments than their counterparts with lower proficiency. For example, Perez-Leroux and Glass (1999), in a translation task eliciting written production by L2ers of Spanish tapping their knowledge of the use of null and overt pronouns, found that the performance of their advanced proficiency learners was better than that of their elementary or intermediate-proficiency informants. This issue is not debated because assuming that L2 proficiency increases with more exposure, input and practice, it is plausible to believe that higher proficiency learners have more knowledge than lower proficiency learners.

Moving to L2 processing, it is equally conceivable to believe that such a difference in the degree of knowledge between lower and higher proficiency learners would be reflected on how they process the target language. Indeed, experimental findings reported in the L2 processing literature support this belief. Results of experiments using neuroimaging techniques reveal differences in the brain activity between lower and higher proficiency learners during on-line sentence processing (see e.g., Wartenburger, Heekeren, Abutalebi, Cappa, Villringer and Perani, 2003; Tatsuno and Sakai, 2005). Also, behavioural tasks recording eye-movements and measuring reading times find different patterns between different proficiency-level learners (see e.g., Frenck-Mestre, 2002 and references cited therein). This proficiency effect has been observed to impinge on processing of different types of linguistic information. Wartenburger et al. (2003) and Frenck-Mestre (2002) reported this effect on the processing of semantic and syntactic information respectively. In the domain of morphosyntax, the type of knowledge of concern to this thesis, sparse evidence exists due to the lack of studies systematically investigating this issue. However, the following study by Rossi, Manfred. Gugler, Angela, Friederici and Hahne (2006) does reveal that processing of morphosyntax is affected by proficiency level.

Rossi et al. (2006) used an electrophysiological measure of event-related brain potentials (ERP), which provides an account of the brain's electrical activity during on-line time course sentence processing, to monitor the brain response to grammatical violations in the aim of investigating the effect of L2 proficiency on speech processing. The authors tested high and low-proficiency L2ers of German and high and low-proficiency L2ers of Italian in a grammaticality judgment experiment conducted during ERP sessions. Informants were placed at Low or High proficiency levels depending on their language-learning history and self-rating on linguistic proficiency as well as their performance on translation tests. This was, as the authors mentioned, due to the lack of standardized language-proficiency tests. In the processing experiment, informants were acoustically presented with correct simple active sentences as well as sentences containing violations of three types, namely a word category violation, a morphosyntactic violation and a combination of the two. In the word-category-violation sentences, a proposition was followed directly by a verb, a position that must be filled by a noun. In the morphosyntactic-violation sentences, a faulty agreement inflection was used on the verb. In the combined-violation sentences, the two types of violations are created. Informants were asked to judge the grammaticality of the presented stimuli. The authors reported that the accuracy rates showed a significant difference between Low and High proficiency learners with the former performing worse than the latter. The ERPs of the Low proficiency and High proficiency groups resulting during grammaticality judgments were compared. What is of interest to us here is the ERPs resulting from grammaticality judgments of sentences containing morphosyntactic violations. It was found that during the judgments of these sentences, there was a brain signal difference between Low and High proficiency learners, as evidenced by (early) left anterior negativity (LAN) effects, assumed to reflect the detection of the violation, observed in the data of the High proficiency groups but not of the Low proficiency groups. Moreover, a centroparietal positivity [P600] effect, which is assumed to reflect a difficulty in the integration of different types of information during reanalysis processes, was observed in both proficiency groups' data but this occurred much later for the low proficiency groups. All in all, these results suggest that L2 proficiency did affect L2ers' processing of morphosyntactic information.

Another possible effect that will be mentioned below is L1 transfer.

2.4.1.2 L1 Transfer

The role of the L1 linguistic system in the acquisition of an L2 is one of the most researched and controversial issues in the field of SLA. A distinction that is commonly made in the L2 literature between two types of L1 transfer is positive transfer and negative transfer (Lado, 1957; Gass and Selinker, 2008). Positive transfer is thought to occur when a certain linguistic property is similar in a learner's L1 and L2, showing up as a facilitation effect on learning that property. Negative transfer, on the other hand, is thought of as a hindrance effect on the learning of a given linguistic property that is different in a learner's L1 and L2. In this sense, difficulties in L2 learning or target-deviant performance experienced by L2ers are attributed to L1-to-L2 negative transfer.

Generative SLA researchers study transfer at the level of mental representations. A number of proposals has been advanced conjecturing on how much of the L1 representations transfer to the L2 and what role they play in building up the L2 representations. These proposals couple with views on the availability of UG in SLA forming together hypotheses on the initial, developmental and end states of SLA. Some proposals were discussed throughout section 2.3 in this thesis. One such proposal is the Full transfer/ Full Access hypothesis (Schwartz & Sprouse, 1994; 1996), which maintains that the L1 grammatical representations transfer in their entirety to L2 and UG remains fully available for L2ers. Another proposal is Organic Grammar (Vainikka & Young-Scholten, 1994, 1996b, 2011, 2013), which proposes that partial transfer of the L1 grammatical system takes place as only lexical categories transfer, with the functional categories getting subsequently built up based on the interaction between a fully present UG and the linguistic input from the target language. There are also other proposals not mentioned in this thesis holding that there is no L1 to L2 transfer. The Initial Hypothesis of Syntax (Platzack, 1996) and the Full Access Hypothesis (Epstein, Flynn & Martohardjono, 1996, 1998) are such proposals (for a review, see e.g., Foley and Flynn 2013).

While SLA research has focused on transfer at the level of representations and how their characterization is affected at a given point or in the course of development, L2 processing, presuming existence of relevant representations already in place, has studied transfer, at the level of processing mechanisms, strategies and routines involved in language production and comprehension. L2 processing might be influenced by transfer of the L1 lexical properties at the surface level or the underlying grammatical categories at a more abstract level and differences between the L1 and the L2 of a learner might "be a barrier to acquiring full nativelike competence and/or fluency in the L2" (Clahsen & Felser, 2006: 4-5). However, evidence in support of transfer at the processing level is inconclusive (Clahsen & Felser, 2006). This inconclusiveness holds true in processing information relevant to different levels of representation, e.g., syntax, semantics and morphosyntax (for a review see e.g., Barto-Sisamout, Nicol, Witzel and Witzel, 2009). As processing of morphosyntax particularly is of interest to this thesis, we expand its discussion hereafter.

Barto-Sisamout et al. (2009: 6-7) propose that mainly four types of L1-L2 relationships hold between components of morphosyntax. These are: (1) a morpho-syntactic feature that is morphologically marked similarly in L1 and L2 (i.e."same/similar"), (2) a morpho-syntactic feature that is morphologically marked in the L2, but not in the L1 (i.e. "L1-L2+"), (3) a morpho-syntactic feature that exists in the L1 and the L2 but under different rules (i.e. "similar but different") and (4) a morpho-syntactic feature that is morphologically marked in the L1 but not in the L2 (i.e. "L1+L2-").

L2 processing research has been examining transfer effects in the relationship types in 1, 2 and 3 and the results seem inconclusive with regard to the role of transfer in the morphosyntactic domain. We saw that the study mentioned above by Rossi et al. (2006) found that the high proficiency (but not low-proficiency) groups of Italian-speaking L2 learners of German and German-speaking L2 learners of Italian were sensitive to subject-agreement violations. In another study by Ojima, Shiro, Nakata, Hiroki, and Kakigi, Ryusuke (2005) testing ERP responses to verb-subject agreement violations by Japanese-speaking high-proficiency L2 learners of English found no P600 effect, indicating insensitivity to these violations. The results of these two studies can be explained resorting to the L1s of the learners as the agreement feature exists in German and Italian but is absent from Japanese. However, the picture is not clear as such. In an ERP study, Frenck-Mestre, Foucart, Carrasco and Herschensohn (2009) tested processing of gender agreement in German-speaking learners of L2 French. Both German and French have gender features but these are morphologically realized differently in the two languages. Test stimuli included gender agreement violations in contexts where the instantiations of the gender feature are different in the L1 and L2 (i.e., agreement between nouns and post-nominal plural adjectives) and in other contexts where the L1 and L2 are similar (i.e., agreement between determiners and nouns and between nouns and pre-nominal plural adjectives). In the ERP results, no P600 effect existed in response to violations in agreement between nouns and post-nominal plural adjectives (L1-, L2+). However, while violations in agreement between determiners and nouns (L1 and L2 similar) elicited P600 effect, the gender violations between nouns and pre-nominal plural adjectives (L1 and L2 similar) did not. Here L1 transfer might explain why the P600 effect did not occur in one context but it cannot explain why it did not occur in another. Bond, Gabriele, Fiorentino and Banon's (2011) study complicates things further. They tested English-speaking low proficiency L2ers of Spanish. Their test stimuli included sentences with violations in number agreement between noun and adjective (feature exists in both L1 and L2 but with different instantiations), in gender agreement between noun and adjective (feature exists in L2 but not L1) and in subject-verb agreement (feature exists in both L1 and L2). The authors found a P600 effect in response to the three types of violations. This inconsistency in the results renders the role of the L1 unclear and an issue in need of further investigation, but at the same time a factor that cannot be neglected when interpreting processing data.

In conclusion, I argued in this section that processing data can be benefited from to resolve controversies over the source of morphological variability. Specifically, it is whether the source of variability is syntactic or not that processing data in this thesis are aimed to resolve. Since other factors are held to affect processing, two such factors that are particularly of interest to this thesis were also discussed. These are L2 proficiency and L1 transfer. It was shown that while the evidence for the former is clear, the evidence for the latter is inconclusive. Both shall be taken into consideration when interpreting the processing data of this thesis' study.

2.5 Conclusion

This chapter has described the conceptual framework for this research on acquisition and processing of L2 morphosyntax. It has discussed the main premises of the Generative Grammar approach to language acquisition, which maintains that human beings are born with an innate ability to learn language. Two classifications in the generative syntactic theory made on syntactic categories (lexical and functional) and linguistic features (interpretable and uninterpretable) has been introduced. The relevance these classifications bear to the SLA of morphosyntax has been discussed. It has been shown that this relevance stems from the proposal that parametric variation across languages is located at features held by functional heads (Chomsky 1993, 1995, 2001), which implies that some features exist in some languages but are absent from others and therefore L2ers might be faced with the task to acquire new features that does not exist in their native language. Indeed, whether L2ers can acquire (uninterpretable) features non-existent in their L1s remains an issue under investigation in the field of SLA.

Discussing SLA of morphosyntax, it has been shown that the observed inaccuracies in the use of morphosyntactic properties by L2ers have led researchers to question the nature of the linguistic representations of these learners. The interpretation to be given to such inaccuracies is what has caused disagreement among researchers. This disagreement appears as a debate on the relationship between overt morphology and their syntactic representations. Two main views have been reviewed in his regard. The first view (i.e., morphology before syntax) maintains that L2ers start their learning task with their interlanguage having only some of the syntactic categories that constitute mature native grammars and they build other categories based on the interaction between the linguistic input and fully available UG. This means that overt morphology is what leads to the instantiation of syntactic categories. Under this view, inaccuracies are seen to be caused by an absence of syntactic categories. On the other hand, the other view (i.e., syntax before morphology) holds that L2ers come to SLA with full syntactic representations in place and their task is restricted to the learning of overt morphology and mapping them into the corresponding underlying representations. Inaccuracies are seen here to be caused by a failure in the mapping procedure.

As morphological variability in the use of tense and verbal agreement morphology is the phenomenon this thesis' study explores and attempts to locate its source, the chapter has reviewed its characteristics based on the findings of previous research. It has been demonstrated that the rates of morphological items suppliance or dropping indicate that this phenomenon is a robustly-observed behaviour in the speech of L2ers. Moreover, it does not only arise in the speech of L2 learners at early to intermediate levels of proficiency, but it also sometimes persists into advanced stages. Although, the learners' L1 does not seem to have an effect on whether they experience this behaviour, it does seem to play a role in the persistence of this behaviour into advanced stages.

Describing the phenomenon of morphological variability, it has been noted that researchers are faced with two main tasks. These are locating the source of morphological variability and the reason for its persistence in the speech of some learners rather than others. The hypotheses reviewed in this chapter address one task or the other. While OG, MSBH and MSIH provide accounts for the variability experienced by all L2 learners at early stages of acquisition, the RDH, PTH and PhRH claim to explain the reason for the persistence of variability.

I argued that understanding how L2ers process morphology could shed light on the nature of the mental representations, which would ultimately help uncover the source of variability. It has been shown that previous research on processing of morphosyntactic properties has revealed differences between adult L2ers and their native counterparts as the former being less sensitive than the latter to morphosyntactic information during speech processing. Two factors that are assumed to generate differences between native and non-native speakers have been discussed. These are L2 proficiency and L1 transfer. While there is clear evidence for the effect of L2 proficiency on the processing of morphosyntactic properties, the evidence for the involvement of L1 transfer in this domain is inconclusive. Despite this, the available evidence necessitates that not only L2 proficiency but also L1 transfer should be taken into consideration when interpreting processing data.

The source of morphological variability and its persistence is far from resolved. Mainly, three different reasons have been proposed by the hypotheses reviewed in this chapter. These are syntactic deficiency (i.e., OG, MSBH, and RDH), processing failure (i.e., MSIH) and phonological deficiency (i.e., PTH and PhRH). The L1 is held to play a prominent role by some of these hypotheses, but thought to have no effect on this phenomenon by the others. While for the RDH it is the absence of certain properties from the L1 that causes their permanent absence in SLA, for the PTH and PhRH it is the L1 phonological constraints on specific structures that cause difficulty in SLA. OG and the MSIH dismiss any role for the L1 in this phenomenon whereas the MSBH proposes that the availability of certain properties in the L1 precipitates their acquisition in SLA. It was noted that the available evidence for these proposals come primarily from production data.

The present study attempts to locate the source of morphological variability and its persistence through testing these proposals against not only production but also perception and processing data. English past tense and verbal agreement are the morphosyntactic properties we seek to explore. These properties, as will be detailed in Chapter three, provide a good testing ground for the syntactic, phonological and processing proposals. Furthermore, the role of the L1 will also be possible to test through including L2ers speakers of L1 Chinese or Arabic. These languages, as will be shown in Chapter three (section 3.3), make it tenable to test both the syntactic and phonological sources of transfer and a careful design of the methodology allows teasing these sources apart.

We turn now to reviewing previous research on the production, perception and processing of past tense and verbal agreement in SLA in the following chapter.

Chapter 3. Production, Perception and Processing Difficulties in SLA of English Past Tense and Verbal Agreement

3.1 Introduction

This chapter will be mainly concerned with reviewing previous research conducted on L2ers' difficulties in the production, perception and processing of English past tense and verbal agreement in order to present the empirical findings that have driven the design of my study. It also describes the morphosyntactic and morphophonological properties of past and verbal agreement in English (the L2 target), Arabic and Chinese (the native languages of research participants) in the aim of providing the linguistic assumptions based on which proposals on the role of the L1 will be tested in the main study of this thesis. The chapter starts with the literature review in section 3.2. The characteristics of the relevant properties in English, Arabic and Chinese are dealt with in section 3.3. Finally, section 3.4 discusses the order of presentation of the following chapters, which present the studies conducted in this thesis.

3.2 Adult SLA of English Past Tense and Verbal Agreement

In this section, previous studies on the production perception and processing (PPP) of tense and verbal agreement will be reviewed. As will be shown below, previous research has focused on the difficulties in either (a) production, (b) perception, or (c) perception and processing of these morphological items, but up to date none has examined the three Ps together in the same study, let alone same learners. For this reason, research on the production on one hand and perception and processing on the other will be reviewed in separate sections in 3.2.1 and 3.2.2 respectively. This will set the empirical background for my study, which uniquely endeavors to examine the three Ps in the same learners. Finally, section 3.3.3 will summarise the key points of the reviewed literature and introduce the present study.

The findings of the reviewed studies will be discussed in the light of the six accounts on the source of morphological variability presented and discussed in the previous chapter (section 2.3.3). This will be in the aim of identifying the research gap and providing rationale for the focus of my study, which seeks to test these accounts in order to locate the source of morphological variability and the reason for its persistence. For ease of reference, the six hypotheses are summarised briefly in Table 3.1 hereafter (see the subsections in 2.3.3 for more details on these hypotheses):

Table 3.1: Summary of Hypotheses' proposals on the source of variability, how it is overcome by L2 learners and the role of the L1 in this phenomenon

Hypotheses	Source of morphological variability	Overcoming morphological variability	Role of L1 in morphological variability
OG	Temporary absence of syntactic representations and/or their features	Building up the corresponding syntactic categories.	None
MSBH	Temporary absence of syntactic representations and/or their features	Building up the corresponding syntactic categories	The presence of certain properties in the L1 speeds up their acquisition in the L2 at relevant points of development.
RDH	Permanent absence of uninterpretable syntactic features	No development takes place	Uninterpretable features not present in L1 are permanently unavailable in adult SLA.
MSIH	Processing failure between overt morphology and underlying syntax	Underlying syntax is complete and development is restricted to learning of forms	Since UG is fully available, presence or absence of properties in the L1 makes no difference
PhRH & PTH	Phonological constraints at the syllable or prosodic levels	Acquiring L2 phonological structures	Phonological structures absent from the L1 cause difficulty in L2

OG = Organic Grammar; MSBH = Modulated Structure Building Hypothesis; RDH = Representational Deficit Hypothesis; MSIH = Missing Surface Inflection Hypothesis; PTH = Prosodic Transfer Hypothesis; PhRH: Phonological Reduction Hypothesis.

We start with production studies in the following section.

3.2.1 Production Difficulties

In this section, the findings of research conducted on variability in the production of past tense and verbal agreement morphology are presented along with the authors' interpretations. Each study will be followed by a brief discussion of its findings in the light of the accounts of variability mentioned above. The discussion will focus on whether the data are relevant to a given hypothesis and if relevant, then how well it accounts for the observed performance. If not, however, I shall suggest what type of data is needed to assess that hypothesis.

Lardiere (1998a, 1998b, 2000, 2003)

In a series of studies (Lardiere 1998a, 1998b, 2000, 2003), Lardiere reported very low production rates of past tense and verbal agreement morphology upon examining longitudinal data collected from a proficient L2 speaker of English. The data from this learner were re-presented in Lardiere (2007). Reviewing these studies here will help clarifying the phenomenon of morphological variability in L2 English although, as we will see shortly, the informant in these studies shows an exceptional performance in the sense that she drops the inflection in considerably higher rates than what is usually reported in other studies.

Lardiere (1998a; 1998b) examined the spontaneous production of a female Chinese-speaking L2 learner of English who had been living in the USA for 10 years when the study commenced. The informant, called Patty, was interviewed three times- once after 10 years of immersion in the USA and the other two times took place after an interval of 8;5 years. Analysing the three data samples, the author found that Patty's production of the inflections on thematic verbs was very low. In contrast, however, Patty's performance on related morphosyntactic phenomena such as nominative case marking in finite past tense contexts and suppliance of correct form of copula and auxiliary *be* that bears an agreement feature was found target-like. Table 3.2 summarises the results of these properties.

Table 3.2: The use of related morphosyntactic properties in the speech of Patty (based on Lardiere, 2007: 74-80)

Sample	Nominative pronouns	Correct form of *be*	verbal agreement marking	Past tense marking
1	49/49 (100%)	57/69 (83%)	2/42 (4.8%)	24/69 (34.8%)
2	378/378 (100%)	50/53 (94%)	0/4 (0%)	191/548 (34.9%)
3	76/76 (100%)	59/63 (94%)	1/22 (4.5%)	46/136 (33.8%)

As can be observed in this table, Patty's nominative case assignment is target-like from the first sample onwards and so is the suppliance of correct form of copula and auxiliary *be* that agrees in number with the subject. This contrasts with simple past tense and verbal agreement marking, which diverges considerably from how native speakers perform and this remains very low throughout the three samples.

Based on these results, Lardiere argues that Patty's syntactic representations for past tense and verbal agreement are intact. For Lardiere, the fact that the inflection is dropped from main verbs does not indicate that the syntactic representations are absent. Auxiliary and copula *be* bear an agreement feature similar to main verbs in 3rd person singular agreement contexts and nominative case is assigned to the subject, as proposed by Chomsky's (1995), by the tense head. Therefore, according to Lardiere, if Patty did not have syntactic categories for tense and agreement, she would not have performed as well as she did on nominative case assignment and suppliance of correct form of *be*.

To account for her informant's performance, Lardiere (1998a; b) offers an explanation similar to that proposed by the MSIH; that is, the absence of the inflection reflects a mapping problem between abstract features and their surface realizations rather than a deficit in the grammatical representations.

Lardiere (1998a; 2003) also observes that Patty's problem with inflections is exacerbated by a phonological difficulty as suggested in the PhRH (see section 2.3.3.6). Lardiere (2003) found that Patty tended to mark past tense on irregular verbs (46.08%)[12] much more than regular verbs (5.80%). Also, the rate of past tense marking in written production[13] (77.92%) was higher than that of oral production (34.47%).[14] Based on this, Lardiere argues that Patty's problem is due, in part, to constraints on the production of the morpheme in consonant cluster codas. Mandarin and Hokkien (the two varieties of Chinese that Patty speaks) disallow consonant clusters and, according to Lardiere, a transfer of such constraint prevents Patty from supplying the inflection when it occurs in such phonological contexts.

Looking at Patty's data in the light of the six accounts of variability (presented in 2.3.3 and summarised here above), we can first observe that Patty's data are irrelevant to OG and the MSBH. This is because OG and the MSBH are hypotheses on the knowledge L2ers start out with and how it develops afterwards but they are not meant to account for persistent problems in SLA. Patty is a proficient speaker of English who has long

[12] Lexical main verbs only.
[13] Written data is taken from 21 email samples collected from Patty over a five-year period.
[14] These rates refer to marking past tense on all verb types.

years of exposure to the target language in an immersion setting and, therefore, having an intact syntactic structure does not contradict OG or the MSBH. These two hypotheses do propose that the absence of the inflection from L2 speech is caused by the absence of its corresponding syntactic categories, which does not seem to be the case for Patty; however, the presence of a stage at which learners drop the inflection even though their grammars include the relevant projection does not preclude the possibility of the existence of a stage at which variability arises because of the absence of syntactic categories. Overall, since OG and the MSBH are hypotheses about the initial state and subsequent stages of development of SLA, Patty's data do not provide testing grounds for these two hypotheses because she is a proficient speaker at the end state of acquisition. Hence, lower L2 proficiency learners are needed to test these accounts. This issue will be taken into consideration in the design of this thesis' study.

As for the RDH, Patty's production of inflectional morphemes seems consistent with this hypothesis, but her performance on related properties does not. According to the RDH, if a specific uninterpretable feature is absent from the L1, it remains absent in adult SLA, which causes persistent difficulties for L2ers. Chinese does not mark verbs for past tense or verbal agreement and thus the underlying syntactic features are assumed to be absent (more on this assumption in 3.3.3). This would lead Chinese L2ers of English to have persistent difficulties in acquiring these properties, according to the RDH. Indeed, Patty's performance does show persistent difficulties with the production of these properties. However, the absence of agreement uninterpretable features would result in the lack of not only the related morphemes on thematic verbs but also agreement marking on copula and auxiliary verbs. In Patty's performance, we have seen that although her production of the inflection is very low, her suppliance of the correct form of *be* that agrees in number with its subject is target-like. This particular finding contradicts a proposal such as the RDH holding that the corresponding syntactic features are absent from the underlying representations. Although a piece of evidence against the RDH exists in Patty's data, this cannot be held conclusive as Hawkins and Lizska's (2003) study (see below) provides evidence to the contrary.

Turning to the MSIH, this hypothesis seems consistent with Patty's performance. Since Patty's performance provides evidence for the availability of tense and agreement syntactic representations, the MSIH seems to be correct in suggesting that the absence of the inflection is caused by a non-syntactic reason. However, based on Patty's data alone, it is premature to conclude that the MSIH is on the right track. What is needed here is checking if the same performance can be attested in other learners from different L1 backgrounds (see White's (2003a) study below).

As for the two phonological hypotheses in this thesis, i.e., PhRH and PTH, Patty's data that are reported by Lardiere in support of the involvement of phonology in explaining morphological variability appear insufficient to conclude that these hypotheses are supported. As shown above, Patty performed better in expressing past tense on irregular verbs than regular ones and in written than oral production. This is indeed the pattern expected if phonological constraints are implicated and, hence, on the face of it the PTH and PhRH seem supported. Nevertheless, this evidence seems insufficient for two reasons. Firstly, although Patty's marking of irregular verbs was higher than that of regular ones, the rate of irregular verb marking was 46.08% and this is lower than what would be expected if dropping the inflection is caused by constraints on the syllable or prosodic structure. Secondly, Patty's better performance in written than oral production might not be a consequence of phonological constraints on her production of the inflection; it is well known that L2ers can control their production better in written than oral tasks as they might have the chance to rely on their metalinguistic knowledge during the former more than the latter types of tasks. These two reasons render the evidence provided in this context insufficient and more data therefore are needed to check the effect of phonology. A more valid assessment of phonological hypotheses should be based on data controlling for the phonological structures that are particularly thought to be problematic (i.e., consonant clusters and prosodic adjunction structures) (more on this below).

White (2003a)

Another study which tested the production of past and agreement morphology in adult SLA of English is White (2003a). This study is particularly interesting because it provides a finding that sheds extra light on Patty's performance. White examined longitudinal spontaneous production data collected from a Turkish-speaking L2 learner of English, called SD. At the time the study commenced, SD had been living in Canada for ten years. White recorded five interviews with SD, with 18 months interval between the first four recordings and the fifth one. Examining the data, White found that "the suppliance of tense and agreement on lexical verbs was quite high (averaging around 80%), noticeably higher than what Lardiere reports for Patty" (ibid: 133).

A comparison between Patty and SD's performance casts doubt on the interpretation given by the MSIH. No difference is expected to arise between L2ers according to this hypothesis. However a considerable difference is observed between Lardiere's and White's informants. It seems therefore that the MSIH alone does not have the power to explain the difference between different-L1 learners in supplying the inflection.

The PhRH, however, might explain the difference between Patty and SD and at the same time save the MSIH interpretation. Indeed the PhRH was advanced to supplement the MSIH. At the level of the syllable structure, Turkish is similar to English and different from Chinese as it allows consonant clusters word finally. Phonological constraints, hence, are expected to affect Patty's, but not SD's, performance. If Patty's performance is a result of phonological constraints, the difference between her and SD cannot be evidence against the MSIH. However, as discussed above, the evidence provided by Lardiere in support of the PhRH is not sufficient to conclude that the phonology is implicated. We will see later in this section if phonological constraints at the level of prosodic structure can explain the difference.

Another possible explanation for the difference between Patty and SD's performance, provided by the RDH, is syntactic. The RDH explains this discrepancy in terms of the absence of the uninterpretable features for past tense and verbal agreement from the L2 grammars of learners who do not have them in their native languages. Given that Chinese does not mark verbs for past and agreement but Turkish does, and thus presumably these features are absent from Chinese but not from Turkish, Turkish learners of L2 English are expected to acquire the features, whereas their Chinese counterparts are not, according to the RDH. This appears to be the attested pattern in Lardiere's and White's studies reviewed here. However, as is shown before, Patty's suppliance of correct form of *be* does not show that the underlying features are absent, which challenges the RDH interpretation.

It can be observed here that the two studies reviewed so far provide data that are either irrelevant to two of the accounts of variability (i.e., OG and the MSBH) or insufficient to assess the other accounts leaving all possibilities open with no weight for one account over another. The following study, however, provides data testing three of these hypotheses (i.e., RDH, MSIH and PhRH) against each other.

Hawkins and Liszka (2003)

Hawkins and Liszka (2003) compared the validity of three competing hypotheses (i.e., RDH, MSIH and PhRH) regarding morphological variability in past tense marking in oral production. Hawkins and Liszka's informants were advanced L2 learners of English from Chinese (n=2), Japanese (n=5) and German (n=5) L1 backgrounds. The L2 proficiency of these informants was measured using a combined grammar/vocabulary test[15] and all of them scored at a rate over 80%. The production of past tense morphology was tested using a film-story telling and past experience recounting tasks.

By comparing their study informants' production of past tense morphology on thematic verbs, Hawkins and Liszka were able to test whether variability was more likely to be caused by a feature-form mapping problem (MSIH), phonological constraints at the syllable structure (PhRH) or a syntactic deficit (RDH). First, if the source of morphological variability resides in the mapping between target-like syntactic structure and surface morphology, the performance of these informants is expected to be similar regardless of their L1 background. Second, if variability is affected by phonological transfer from the L1 in line with the PhRH, the problem is expected to be more marked in the Chinese and Japanese than German informants. This is because German is similar to English as it allows word-final consonant clusters, but both Chinese and Japanese disallow them. Third, if the source of variability is L1-tansfered constraints at the syntactic constraints, the Japanese and German learners should perform similarly in a target-like manner contrary to the Chinese learners, who are expected to perform below the other two groups. This is because German and Japanese mark past tense on verbs, different from Chinese, which does not.

The results of Hawkins and Liszka's study are presented in Table 3.3. It was found that the Chinese participants performed significantly below their Japanese and German counterparts as they were less likely to mark past tense on regular and irregular verbs. Japanese and German informants patterned together showing similar rates in their suppliance of the morphology.

[15] The Oxford Placement Test (Allan, 1992) and the Vocabulary Level Test (Nation, 1990)

Table 3.3: Production of regular and irregular past by Chinese, Japanese or German learners of L2 English (based on Hawkins and Liszka, 2003)

Verb Type	Chinese	Japanese	German
Regular	25/40 (62.5%)	137/149 (91.9%)	52/54 (96.3%)
Irregular	64/76 (84.2%)	252/270 (93.3%)	79/83 (95.2%)

As can be seen here, these results meet the predictions of neither the MSIH nor the PhRH. Hawkins and Liszka interpreted these results as evidence against these views of the source of L2 speakers' variability. The observed pattern, however, is consistent with the predictions of the RDH.

Hawkins and Liszka further observed that the Chinese participants retained final –t/-d with monomorphemes (82%) more often than with regular past tense verbs (62.5%). This is interpreted by authors as an indication that the problem their Chinese informants had was not phonological.

Hawkins and Liszka maintain that although their account explains the variability experienced by the Chinese informants, it still needs to explain why the rate of suppliance of past tense marking is 62% rather than 0% as would be expected if the syntactic feature is absent. To account for this, Hawkins and Liszka (2003: 38) suggest that "linguistic theory appears to need to allow operations which apply to strings post-syntactically, that is in the morphological component or following vocabulary insertion". The task of this 'post-syntactic operation' is assumed to be monitoring the linguistic output and checking it before spell-out. This output checking process causes the insertion of the inflected form of the verb (V-ed) when the Chinese speakers are able to detect or monitor the ambient discourse for 'pastness'. As a result, the work of this 'monitor' depends on the context of the speech and thus gives rise to random use of inflected and uninflected verb forms.

Although the findings of Hawkins and Liszka’s study are suggestive, they should be interpreted with caution as the authors themselves warn. This is because the number of participants is very small and thus further research is needed to check if these findings are generalisable.

The studies reviewed in the rest of this section puts more focus on exploring the role of phonological constraints on morphological variability.

Goad, White and Steele (2003) & Goad and White (2006)

Goad, White and Steele (2003) explored the possibility of the involvement of L1 prosodic constraints in the difficulties experienced by L2ers in the production of verbal agreement and past tense morphology. The study participants were 12 Mandarin-speaking L2ers of English. Their L2 proficiency was measured by means of the grammar and vocabulary sections of the English Language Institute placement test and they scored at the high intermediate / low advanced range. At the time of testing, the informants had lived in Canada for a period ranged from 6 months to 5 years.

The participants were tested on their knowledge of tense and agreement morphology by a grammaticality judgment task and all of them performed in a native-like manner. This indicated, according to Goad et al., that those learners represented the tense and agreement properties in their interlanguage grammar, at the metalinguistic level at least.

Then, oral production data were collected using a picture description task in which the participants had to describe sets of pictures. This task targeted the elicitation of tense and agreement related properties. The data revealed the incidence of properties indicating the presence of tense and agreement representations in those learners' grammars. First, assignment of nominative case on subjects, a property associated with the tense head, was perfect (at a rate of 100%). Secondly, the incidence of copula (be*)* and auxiliaries (be, have and do), which carry tense and agreement features, was very high at a rate of 97% and 87% respectively. In contrast, it was found that past tense and verbal agreement marking on lexical verbs was omitted to a great extent. Scores and rates of supplying past and verbal agreement marking on lexical verbs are presented in Table 3.4

Table 3.4: Production of regular and irregular past and 3rd person agreement by Chinese-speaking L2 learners of English (based on Goad et al., 2003: 255)

	Score	%
Regular past tense	16/28	57
Irregular past tense	55/71	78
3rd person agreement	57/201	28

As can be seen here, low rates of suppliance are observed, especially for regular past and 3rd person agreement marking.

To explain these results, Goad et al. propose the PTH introduced in 2.3.3.5, which postulates that a transfer of L1 prosodic constraints limits L2ers' production of the inflections. Specifically, problems are expected to arise in L2 production of inflectional morphology when the L1 and L2 of learners differ in how they prosodify these items. As mentioned before, the authors propose that the English past *–ed* and 3rd person agreement *–s* are attached externally to the PWd creating an adjunction structure that violates EXHAUST and NONREC constraints simultaneously (see also section, 3.3.1 below). By contrast, this structure is claimed to be not present in Chinese. Chinese has an aspect marker, i.e., /-*l*/, but Goad et al. argue that this suffix is attached to the Foot level in the prosodic structure (word-internally). Hence, while the inflection in Chinese is adjoined internally to the word at the foot level, the inflection in English, in contrast, is adjoined externally to the word at the higher PWd level. This mismatch in the prosodic structure between English and Chinese and the unavailability of adjunction structure in Chinese language is claimed to be the source of difficulty for Chinese L2ers of English in supplying tense and agreement inflections.

According to Goad et al., the PTH predicts two patterns of behavior in case of Chinese learners of English. The first pattern is that some learners will delete the inflection comprehensively due to their realization of the mismatch between what their grammar permits and the prosodic structure needed for producing the English inflection. The second pattern is that other learners will use the inflection variably because the inflection will be supplied when it can be accommodated word internally similar to the Mandarin aspect suffix (Goad et al., 2003: 254).

Goad et al.'s study data contained more contexts for 3rd person *–s* than the past tense *–ed* and, accordingly, they elaborated more on the analysis of the verbal agreement data. The predicted two patterns in behavior were borne out as six participants deleted the inflection comprehensively and their proportion of suppliance was 10% and the other six participants deleted the inflection variably supplying it at a rate of 48%.

Most importantly, Goad et al. propose that variability can also be predicted by the PTH. They argue that there are three conditions under which the English inflection can be prosodified word internally and, thus, causes no difficulty. These are as follows

1. Inflection as onset: the inflection is followed by a word that starts with a vowel so the inflection can be syllabified as an onset as in "builds on" [bɪldzan].
2. Inflection as coda: the inflection is added to a base ending with a sibilant so a schwa epenthesis occurs and it can by syllabified as a coda as in "races" [rejsəz].
3. Inflection as foot-internal: the inflection is added to a base which is ….VX] in shape (X = V or C) so it can be prosodified to the foot as in "fills" [fɪlz] or "sews" [sowz].

(Goad et al., 2003: 257)

In these three conditions the inflection should be supplied because it can be prosodified without violating the L1 constraints of the informants (i.e., Mandarin). This contrasts with cases where the inflection cannot receive a stem-internal analysis such as "with bases which are …VXC] in shape as in, for example, [bɪldz] 'builds', [kijps] 'keeps' (before a consonant-initial word or pause)" (Goad et al., 2003: 257). The group results of the 6 participants who supplied the inflection variability are presented in Table 3.5.

Table 3.5: Production of 3rd person –*s* by stem shape by six Chinese learners of L2 English (Based on Goad et al. 2003: 258)

	Percentage
Inflection as onset	75%
Inflection as coda	27%
inflection as foot-internal	68%
no option for inflection inside PWd	9%

Goad et al. take these results to be consistent with the PTH. However, as can be seen in this table, the production rate of the inflection where it is syllabified as coda is very low and thus not compatible with what is expected by the PTH. The authors explain this as a result of another phonological constraint transferred from the L1; this is thought to be a constraint against obstruent codas in unstressed syllables

In another study investigating the same phenomenon, Goad and White (2006) presents results showing that these constraints can be overcome by Mandarin speakers and, thus, target-like prosodic representations are attainable. The study is presented hereafter.

Goad & White (2006) examined the production of past tense marking by 10 intermediate level Mandarin-speaking learners of English. In a sentence completion task, participants were presented with the beginning of a sentence followed by two endings, viewed on a computer screen; they were given 12 seconds to choose and memorise one ending and then say it aloud after it had disappeared from the screen. The following is an example from the study (from Goad and White, 2006: 252):

(3.1) Last night after dinner…

- you show me photos of your daughter
- you showed me photos of your daughter

The results showed that the participants in the study supplied past tense morphology above 90% of the time. This finding is in contrast with the results of Goad et al.'s (2003) study, in which it was found that high intermediate / low advanced-level Mandarin-speaking L2ers of English were unable to supply the morpheme in contexts requiring the adjunction structure. Goad and White (2006) explain the difference in the results of the two studies as an effect of the task used in eliciting the data; while Goad et al. used a picture description task in which the participants' attention was not drawn to the focus of the study, Goad and White used a sentence completion task in which participants were given two possible endings making them more aware of the focus of the experiment. Nevertheless, the informants' performance in the latter study was interpreted by the authors as indication of success in acquiring the English adjunction structure.

This study shows, according to Goad and White, that building target-like prosodic representations for the English verbal inflection is possible through minimal adaptations from the L1. Goad and White conjecture that the structures that could be adapted from the L1 (Mandarin in this case) to create adjunction structures needed for accommodating the English infection are "PWd dominating PWd (PWd–PWd), required for lexical compounds" and "PWd directly dominating **σ** (PWd– **σ**) at the right edge, the structure required to prosodify three-syllable PWds which contain only one foot" (Goad and White, 2006: 251).

To sum up, Goad et al. (2003) observed that the English inflection was supplied by their Mandarin-speaking informants on verbs that have, or occur in, specific prosodic structures but not others. They proposed that L2 production is controlled by prosodic L1-transferred constraints. Persistent difficulty in supplying the L2 inflection arises when the L1 and L2 mismatch in how they prosodify these items. In Goad and White (2006), Mandarin-speaking informants supplied the L2 inflection in a target-like manner. This led the authors to propose that target-like prosodic representations are ultimately attainable through resorting to other structures in the native language and adapting them to accommodate the L2 inflection.

Based on these two studies, two important issues should be taken into consideration to check the validity of the PTH in explaining the L2ers' difficulty with producing the inflection. These are (1) the L2ers' native language and (2) the L2ers' proficiency level. First, this hypothesis is built on observations on the performance of native speakers of Mandarin, a language that prosodify the inflection in different manner from English. To test this hypothesis, learners from a language that prosodify the inflection similarly to English should also be tested to check if the observed performance is a result of L1-transfered constraints or a general mechanism. Second, although prosodic constraints cause problems in producing inflectional morphology, this is deemed temporary and it is not clear at what point of linguistic development these can be overcome. Goad et al.'s (2003) informants were in the high-intermediate/ low advanced range of proficiency and they had not had acquired the required prosodic representations. Goad and White's (2006) informants, by contrast, were lower in proficiency (i.e., intermediate), but they produced the inflection in a target-like manner and, according to the authors, had acquired target-like prosodic representations. Thus, testing learners at a given proficiency level and not finding support for the PTH might be merely because the learners have already acquired the intended structure. One way to overcome this problem is through testing learners at different proficiency levels. Therefore, to test the PTH, L2ers from different L1 backgrounds and at a range of L2 proficiency levels are needed. The following study by Campos Dintrans (2011) tested the PTH through including L2ers speakers of different L1 backgrounds.

Campos Dintrans (2011)

Campos Dintrans (2011) explored the PTH and RDH interpretations' validity in explaining morphological variability through investigating the production of past tense and plural marking in L2 English by native speakers of Mandarin (n= 15), Spanish (n= 13) and Japanese (n= 11). The L2 proficiency of these informants was assessed by means of the multiple-choice grammar test and the multiple-choice vocabulary test of the Michigan Placement Test. All participants were at the same level and had a combined score of around or above 75%.

The participants' production of past tense was tested in three different tasks. The first was a picture description task, in which participants were shown a set of pictures with some prompts and they were asked to describe them. The other two tasks were an identical version of a sentence completion task but completed by participants once manually (written production) and the other time orally (oral production). This written/oral completion test comprised of sentences with missing verbs and with each sentence three non-inflected verbs were provided. In the written task, participants had to choose one verb and write it down in the correct form. In the oral task, they were asked to read the sentences aloud and fill in the gap with the appropriate form of one of the given verbs. (3.2) is an example from the written/oral completion test.

(3.2)
(a) Johnny had a terrible headache, so he _______ a glass with water and took two aspirins.

[fill write type]

The choice of the L1s (i.e., Mandarin, Spanish and Japanese) and the different types of data collected (i.e., Oral and written production) allowed the author to test the PTH and RDH against each other as they have different predictions on the performance of the informants of the study. Firstly, the RDH predicts that the Spanish and Japanese participants will perform better than the Chinese participants because Spanish and Japanese languages, contrary to Chinese, are assumed to have the uninterpretable feature for past tense. By contrast, the PTH predicts that there would be no difference in the performance of the three groups of learners in the oral production because, according to Campos Dintrans, their native languages similarly prosodify the inflection word internally, different from English in which the inflection is prosodified word externally. Secondly, according to Campos Dintrans, while the RDH expects no difference to arise between the oral and written production, the PTH predicts that the written production should show higher rates of suppliance of the past tense inflection than the oral production. The following table (3.6) summarises the results of non-native groups in addition to a group of native speakers tested in the same study.

Table 3.6: Production of English past tense morphology by native and non-native speakers in three different tasks (Based on Campos Dintrans, 2011: 121-199)

	NS	Spanish	Mandarin	Japanese
Sentence Completion: Oral	96.78%	83.24%	79.49%	88.41%
Sentence Completion: Written	98.98%	97.95%	97.5%	94.08%
Picture description: Oral	100%	86.9%	87.6%	90.9%

Campos Dintrans reports that statistical tests performed on results of each task separately revealed no difference among the non-native groups, but the suppliance of the inflection was significantly higher in the written than the oral production for both Mandarin and Spanish, but not Japanese, groups. According to Campos Dintrans, these results are more compatible with the PTH.

Based on the finding that oral production was markedly lower than written production, Campos Dintrans concludes that morphological variability is more likely to be due to phonological constraints. However, whether these phonological constraints are at the syllable level in line with the PhRH or at the prosodic level as posited by the PTH is not clear yet. The author, therefore, proceeds to test this as follows: (1) comparing rates of inflection suppliance in consonant cluster contexts with rates of target realisation of consonant clusters in monomorphemic words and (2) comparing the suppliance rates of past tense inflection in consonant cluster and non-consonant cluster contexts. The reason for such comparisons is to examine if the inflection production is affected by a ban on consonant clusters word finally, which would lead learners to reduce consonant clusters not only in inflected verbs but also in monomorphemic words. If this is not found to be the case however the prosodic constrains would be the source of the problem according to the author. The results of these comparisons are presented in Table 3.7.

Table 3.7: Oral production of regular past tense *–ed* in consonant cluster and non-consonant cluster codas compared to the production of consonant cluster codas in monomorphemic words by native and non-native speakers of English (Based on Campos Dintrans, 2011: 143-157)

	NS	Spanish	Mandarin	Japanese
monomorphemes with CC cluster	98.22%	74.79%	87.7%	94.54%
simple past tense forms with CC cluster	97.33%	76.63%	71.49%	83.82%
simple past single C	97.77%	92.94%	91.33%	93.93%

Campos Dintrans looks at the results presented in this table based on the finding that the written production of past tense was target like and higher than the oral production of the same morpheme, which was taken as indication that the underlying syntactic representation for past was acquired by all learners. In this sense, difficulties in supplying the inflection are claimed to be due to phonological constraints either at the syllable level or the prosodic level. Campos Dintrans interprets the results in Table 3.7 as follows. For Japanese learners, their results show that they have acquired the syllable and prosodic structures required for producing past tense as no significant difference was detected between the production of the consonant clusters in inflected verbs and monomorphemic words or the consonant clusters and non-consonant clusters in inflected verbs. For the Mandarin speakers, the difficulty is a result of a blend of phonological constraints at the prosodic level (as exhibited by the lower rates of CC production in inflected verbs than in monomorphemic words) and at the syllable level (as exhibited by the higher rates of inflection suppliance in single consonant codas than consonant cluster codas). Finally, for the Spanish speakers, while they appear to have acquired the target prosodic representations (as indicated by the similar rates of consonant cluster production in inflected verbs and monomorphemic words), they still have difficulty with phonology at the syllable level (as indicated by both the similar rates of consonant cluster production in inflected verbs and monomorphemic words and the higher suppliance of the inflection in single consonant codas than consonant cluster codas).

Campos Dintrans argues that the results of his study provide evidence in support of the PTH and against the RDH. The argument is mainly based on the better performance of all non-native groups in the written task compared to the oral tasks. However, is phonology the only possible source of such difference? This difference might also be attributed to the learners' ability to control their production more in written than oral tasks. Moreover, although the comparisons between the phonological structures presented in Table 3.7 can work as a test for the presence or absence of phonological constraints at the syllable level, they cannot be taken as convincing evidence in support or against phonological constraints at the prosodic level. In a domain such as the production of inflectional morphology in which many factors are believed to cause difficulties, we expect the evidence for or against the involvement of prosodic constraints to come from the analysis of and comparisons between the prosodic structures that the inflection creates when added to verbs just as Goad et al. (2003) did. This does not mean that other means for testing the hypothesis are invalid; rather the intention is that a direct testing provides more convincing evidence. We take this into consideration in this thesis' study through analysing the specific prosodic structures mentioned by Goad et al. (2003).

To recap, this section has reviewed previous research conducted on the variability that L2ers experience in producing English past tense and verbal agreement morphology and discussed its findings in the light of six hypotheses on the source of this phenomenon in the aim of identifying the research gap, which the research at hand attempts to bridge. It has been observed that morphological variability in L2 English persists in some learners' speech but not others. The L2ers' native language has been identified as a possible source of the problem. Lardiere (1998a, 1998b, 2000, 2003) investigated the phenomenon in a Chinese speaker who was thought to be at the end-state of SLA of English. She showed that the difficulty experienced by her informant was only in producing the inflection but not in other related morphosyntactic properties, which, according to the author, dispelled the possibility that the source of her informant's problem was syntactic. Lardiere proposed that a mapping problem between the underlying representation and surface forms exacerbated by L1-transferred phonological constraints at the syllable level could explain this difficulty. White (2003a) reported consistent production of the inflection by a Turkish L2er of English. A comparison between Lardiere's and White's informants posed the L1 as a possible source of the phenomenon. Investigating this possibility, Hawkins and Liszka (2003) looked at the phenomenon in advanced L2ers of English native speakers of Chinese, Japanese or German. The findings of Hawkins and Liszka's study demonstrated that L1 syntactic transfer rather that L1-transferred phonological constraints at the syllable structure was more likely to account for their informants' performance. The possibility of L1-transferred phonological constraints in prosodic structure was investigated by Goad, White and Steele (2003) and Goad and White (2006), which confirmed the involvement of such constraints in Chinese L2ers of English production of the inflection. The final study the review by Campos Dintrans (2011) examined the phenomenon in Mandarin, Japanese or Spanish speakers advanced L2ers of English. The author argued that a combination of L1-transferred phonological constraints at the syllable and prosodic structures could account for his informants' performance better than L1-transferred syntactic constraints. Finally, discussing these studies, I argued that they do not provide either testing grounds for some proposal on the source of morphological variability at early stages of L2 linguistic development (i.e., OG and MSBH) or conclusive evidence for or against any of the other proposals discussed here (i.e., MSIH, RDH, PTH and PhRH).

The review in this section has focused on studies that tracked the source of morphological variability in the production of L2ers of English past tense and verbal agreement morphology. The next section reviews studies that have explored the perception and processing of the same properties.

3.2.2 Perception and Processing Difficulties

We have seen that the debate on the source of morphological variability in generative SLA is a question of whether this phenomenon is caused by a (temporary or permanent) syntactic deficit (OG, MSBH and RDH) or non-syntactic deficiencies such as a processing failure in production (MSIH) or a phonological difficulty (PTH & PhRH). In the previous section, studies that tried to resolve the debates through testing the proposals against production data were reviewed. The production studies however have provided inconclusive evidence, fanning the debates rather than resolving them. This section reviews studies that explored other types of data, namely perception and processing in the aim of locating the source of adult L2ers' difficulty with past and verbal agreement morphology.

It is worth clarifying here what I mean by the perception of past and agreement morphology. Tatham and Morton (2006: 194) explain that speech perception has two aspects, which are 1- "the acoustic signal and the way the listener access this signal using hearing" and 2- "the interpretation of what the listener has heard". I use the word perception here to denote both aspects. Therefore, the perception of past tense and verbal agreement morphology includes both of the phonological decoding of the inflection as V+ed/ V+s and interpreting it through assigning a meaning or function. Relating this to syntax, it can be said that while a failure in phonological decoding does not necessarily entail the existence of non-target like syntactic representations, a failure in assigning a meaning or function might be caused by either a failure in phonological decoding or the existence of non-target like representations.

Data on the perception and processing of the properties under scrutiny could help reveal whether the source of difficulty is syntactic or non-syntactic. It is conceivable that an impairment in or absence of the representation would affect the PPP similarly based on the assumption that the three modalities share the same syntactic representations. Therefore, if the source of morphological variability in the production is syntactic, successful perception and processing are not expected to occur, but if the source is non-syntactic, difficulty in production is not expected to extend to perception and processing. However, the issue is not as straightforward as it seems; difficulty in perception and processing might not only be a reflex of impaired or absent syntactic representations but also L2ers' poor phonological decoding abilities might generate similar difficulties (McDonald and Roussel 2010). What makes this latter possibility further tenable is the phonological environment in which regular past and verbal agreement are realised. As will be elaborated in section 3.3.1 below, the addition of past *–ed* and agreement *–s* to verbs creates four different phonological context among which a cluster of consonants is in coda position (e.g., walked /kt/, walks /ks/). Such clusters are thought to cause difficulties not only in production but possibly in perception especially for learners whose native languages disallow them (Lardiere, 2003). Indeed, a study by Solt, Pugach, Klein, Adams, Stoyneshka and Rose (2004) found that the regular past inflection in complex codas poses a perceptual challenge for L2ers (more on this below). Therefore, perceptual difficulties should not be taken as a reflection of non-target like syntactic representation as they could also be caused by a phonological decoding difficulty.

The number of studies that has investigated the perception and/or processing of past and verbal agreement by adult L2ers is relatively small (Johnson and Newport, 1989; McDonald, 2000, 2006; Solt et al., 2004; Chen, Shu, Liu, Zhao and Li, 2007, Sato and Felser, 2008; McDonald and Roussel, 2010). All of these studies (apart from Solt et al., 2004) focused on learners' sensitivity to violations by using grammaticality judgement tasks (GJT). In such a task learners are presented with stimuli involving correct use of the morphology or violations of its use and asked to judge whether they are grammatical or ungrammatical. The learners' behavioural responses (the accuracy in the decision on the grammaticality of the stimuli) as well as RTs (the speed of their decision) and ERPs (the brain signals at the time of making the decision) are taken as indication of (un)successful perception and processing. We shall start with Johnson and Newport

(1989) which has been taken as a benchmark by many in the general conceptualisation of SLA theories.

Johnson & Newport (1989)

Johnson and Newport (1989) is one of the early studies that examined the receptive knowledge of past and verbal agreement morphology in addition to ten other grammar rules of English. The participants of their study were 46 native Chinese or Korean speakers, who had had at least five years of exposure to English in the USA, in addition to a control group of 23 native speakers. Since the main purpose of Johnson and Newport's research was to investigate age effects in SLA, the non-native speakers were divided into two main groups according to their age of arrival to the USA – early arrivals (those who arrived before the age of 15, n=23) and late arrivals (those who arrived after the age of 17, n=23). The early arrivals group was also further divided into three age groups – 3-to-7 group, 8-to-10 group and 11-to-15 group – and the late arrivals were divided into two groups – 17-to-24 group and 25-to-39 group. A GJT was used in the study to test learners' knowledge of 12 rule types of English grammar including wh-questions and word order as well as past tense and verbal agreement morphology (for details, see Johnson & Newport, 1989: 68-77). The GJT consisted of 276 sentences (136 grammatical and 140 ungrammatical). With regard to the past and verbal agreement morphology, eight pairs of sentences (half grammatical and half ungrammatical) were constructed. While the grammatical sentences contained the target morpheme in an obligatory context, the ungrammatical sentences were similar in every respect except for containing violations in one of four formats: (1) morpheme is omitted from obligatory context, (2) inappropriate morpheme is used, (3) irregular verb inflected regularly (past verbs only) or (4) regular inflection attached to already irregularly marked verb (past verbs only). Johnson and Newport (1989: 86) argue that only ungrammatical sentences can be considered to be testing a specific grammar rule. Their rationale is that when a learner judges an ungrammatical sentence incorrectly, it can be said that s/he has problems with that specific structure creating the ungrammaticality but, in contrast, when the learner judges a grammatical sentence incorrectly, it is not clear which part of the sentence is problematic for him/her. The stimuli were presented aurally to participants and they were asked to give a response on whether the presented sentences were grammatical or ungrammatical.

The most well known finding of Johnson and Newport's study is that there was a strong linear relationship between age of exposure to English and the performance in the GJT with the learners who arrived between the age of 3 and 7 performing similar to native speakers and all other learners who arrived after the age of 7 performing significantly below. The authors interpreted this as an effect of a critical period on language acquisition. Of more interest to this thesis are the findings that are related to past and verbal agreement. It was found that the proportion of errors in judging ungrammatical sentences was around 35% in past tense items and above 20% in verbal agreement items for late arrivals and below 20% for past tense items and below 15% for verbal agreement items for early arrivals.[16]

The results of the late arrivals are particularly interesting because they show that difficulty with morphology extends to perception, giving prima facie support to syntactic accounts of morphological variability. If those learners had the relevant grammatical knowledge, the violations of the rules were expected to prompt a strong sense of ungrammaticality and lead them to judge the sentences consistently just as native speakers did. The average number of years of exposure to English in the USA for those learners was 9,9 years (Johnson and Newport, 1989: 69); despite this long exposure, their knowledge differed significantly from that of native speakers. Therefore, their inability to consistently detect the violations might be suggestive of their lack of the underlying grammatical knowledge.

Another study that has detected similar difficulties in speakers of different L1 backgrounds is McDonald (2000). This is presented below.

McDonald (2000)

McDonald (2000) is another study that explored the receptive knowledge of past and verbal agreement in L2 English but with different L1 groups this time. The purpose of this study was also to investigate age effects in SLA and it used an auditory GJT. The participants were 28 native speakers of Spanish and 24 native speakers of Vietnamese in addition to 14 native speakers of English as a control group. The Spanish participants were divided into early learners (those who arrived to the USA before the age of 5, n=14) and late learners (those who arrived to the USA after the age of 14, n=14) and the Vietnamese participants were divided into early learners (those who arrived to the USA before the age of 5, n=14) and child learners (those who arrived to the USA between the

[16] See Johnson and Newport (1989, Fig 3: 87).

age of 6 and 10, n=10). The GJT used in this study was modelled on Johnson and Newport's (1989) test examining the same 12 rules of English grammar including past tense (11 grammatical and ungrammatical pairs) and verbal agreement (5 grammatical and ungrammatical pairs). Both behavioural response (accuracy) and RTs were recorded in McDonald's study.

The general finding of McDonald's study was similar to that of Johnson and Newport, that is, the ultimate performance was predicted by age of arrival as that the earlier the better. However, she also found that the native language of learners was another predictor of better performance since the Spanish early arrivals (with a language similar to English in many respects) performed similar to English natives, while the Vietnamese early and child arrivals (with a language different from English) lagged significantly behind. What is relevant to us here is the accuracy and RTs in past and verbal agreement items. These are summarised in Table 3.8:

Table 3.8: Percent mean accuracy % and mean RTs in millisecond (in parentheses) by English natives and Spanish and Vietnamese learners of English (Based on McDonald 2000, Tables 2, 3, 5 and 6: 405-412)

	Past tense		Verbal agreement	
English natives	95.3	(2.700)	96.4	(2.158)
Spanish early learners	87.7	(2.818)	82.1	(2.270)
Spanish late learners	75.3	(3.562)	44.6	(3.365)
Vietnamese early learners	76.6	(3.240)	69.6	(3.058)
Vietnamese child learners	72.7	(3.200)	45.0	(2.958)

McDonald reports that the statistical tests showed that only Spanish early learners performed similar to native speakers with no significant differences in accuracy or RTs but all other groups were significantly different from native speakers in both of accuracy and RTs. These results clearly show the difficulty L2ers had in detecting the violation.

We assume that if the source of variability is syntactic, it would be mirrored as a difficulty in perception and processing. McDonald's study results show that the difficulty with morphology is not only a production problem but also a perception as well as processing problem. However, these results do not inform us much about the validity of the syntactic accounts under testing in this thesis (i.e., OG, MSBH and RDH). This is because with a length of exposure of at least 3 years for Spanish late arrivals, 9 years for Vietnamese child acquirers and 15 years of Vietnamese early arrivals, these participants are very likely to have passed the stage at which the OG and MSBH claim that they lack grammatical representations. Also, according to the RDH, neither the Spanish nor the Vietnamese informants are assumed to lack the syntactic representations for past and agreement. The reason for this is that Spanish has both properties and the Vietnamese participants started learning English before puberty and thus no critical period effects are expected. Therefore, further research controlling for the L1 and age of first exposure is needed in order to test these syntactic accounts.

Furthermore, one methodological feature of Johnson and Newport's (1989) and McDonald's (2000) studies renders the syntactic account not the only possible interpretation. This methodological feature is the auditory presentation of the stimuli. Johnson (1992 cited in Jiang 2004: 608) points out that not only grammatical competence is needed to complete the auditory GJT successfully but also phonological decoding abilities become critical in such a task. L2ers might have poor phonological decoding abilities that lead them to perform inconsistently in aurally presented stimuli. This is a competing interpretation to the syntactic account and it also coincides with non-syntactic accounts of variability (e.g., MSIH). The phonological decoding abilities along with other non-syntactic factors affecting L2er's performance are investigated empirically by McDonald (2006) and McDonald and Roussel (2010) presented below.

McDonald (2006) & McDonald and Roussel (2010)

McDonald (2006) explored the possibility that L2 late learners' poor performance is caused by "processing difficulties due to (1) low L2 working memory capacity, (2) poor L2 decoding, and/or (3) inadequate L2 processing speed" (ibid: 381). The informants of the study were a group of 50 late learners from various L1 backgrounds (i.e., Berber, Bosnian, Chinese, German, Hind, Japanese, Maranthi, Portuguese, Romanian, Spanish, Tagalog, Turkish, Serbo-Croatian, Slovenian and Vietnamese). In addition to the main task of the study, which was a GJT, three other tasks were used to measure the participants' working memory capacity (size judgment task), decoding ability (gating task) and processing speed (word detection task).[17]

[17] In the size judgment task, the participants were given a set of words and asked to retell them in the order of the size of the referent. For example, the participants heard "goldfish, pig, needle" and they had to repeat them in the order from smaller to bigger, i.e., "needle, goldfish, pig". In the gating task, words were divided into segments and the first segment is presented to the learner to guess the word; if the word was not guessed, the second segment was given and so on until the word was guessed. In the word

The results of this study revealed that the non-native speakers performed significantly below the native speakers in all tasks. More importantly, it was found that the accuracy results and the RTs in the GJT correlated significantly with the working memory capacity and decoding ability for the non-native speakers, indicating that these factors could explain the participants' poor performance in the GJT.

Furthermore, in another experiment in McDonald's study, the author placed a group of native speakers under stressors[18] during performing the GJT and she found that they showed a performance that was parallel to that of non-native speakers in the first experiment. Based on these findings, McDonald concluded that the poor performance of non-native speakers in the GJT can be better accounted for by limitations of their working memory capacity and decoding abilities.

In another similar study, McDonald and Roussel (2010) tested other possible performance-affecting factors that are relevant to the inconsistent judgment of specifically past tense items. Motivated by findings of previous research showing that L2 non-native speakers differ from native speakers in their productive and perceptual phonological abilities and in lexical access and knowledge, McDonald and Roussel explored the possibility that the poor performance in response to items testing past tense morphology in GJT is caused by difficulties in (1) the perception of the phonological structure in case of regular past and (2) access to relevant lexical knowledge in case of irregular past. The participants of this study were 23 non-native speakers who came from a variety of L1 backgrounds (i.e., Spanish, Romanian, Portuguese, French, Russian, Dutch, Arabic, Swahili and Vietnamese) and 15 native speakers of English who served as a control group. At the time the study commenced, the non-native speakers were residents in the USA for a length average 3.2 years (ranging from one month to 17 years). A self rating measure of proficiency was introduced to the non-native participants, and on a five-point Likert scale (1 poor and 5 excellent), the ratings' average was 3.74.

---

detection task, the participants were presented aurally with word followed by a sentence and they were asked to press a button as soon as they heard that specific word in the sentence.

[18] The participants were either asked to remember a number of 4 or 7 digits in length during judging the stimuli, put under pressure to respond, presented with the stimuli under noise or presented with the stimuli at twice the rate of speech (for details, see McDonald 2006: 392-393).

McDonald and Roussel assessed their participants' phonological abilities through a gating task (see footnote 17, above) and a word pair discrimination task that focused on their ability to detect [t] and [d] phoneme in coda position. In the word pair discrimination task, the participants were aurally presented with similar word pairs (e.g., coal- coal/ cold-cold) or word pairs that differed only in the final phoneme (e.g., coal-cold) and they were asked to judge whether the pairs were similar or different. Then, the participants' lexical knowledge and access was assessed by their response and RTs in a picture naming task, in which they were presented with pictures depicting verbs and asked to name them. A third task in this study tested the participant's knowledge of past tense through giving them bare verbs and asking them to say their simple past tense form aloud. Finally, the GJT focused on testing past tense morphology, so it consisted of grammatical and ungrammatical sentences differing only in whether the past form of the verb was correct or incorrect.

The results of McDonald and Roussel's study showed that the non-native speakers performed significantly below the native speakers in both production and grammaticality judgments of past tense. Most importantly, a strong correlation was detected between non-native participants' phonological abilities and lexical access/ knowledge, and their production and judgment of regular and irregular past, respectively. These led the authors to conclude that late learners' poor performance is explainable by limitations on their phonological decoding abilities and lexical knowledge and access.

All in all, what these two studies by McDonald (2006) & McDonald and Roussel (2010) reveal is that lack of grammatical knowledge is not the only interpretation for L2ers perceptual difficulties because other factors such as phonological decoding abilities, working memory and lexical access and knowledge can have similar effects. Therefore, these factors, if not controlled in the task design, should be taken into consideration in interpreting L2ers' performance not only in GJTs but also in other tasks in which the stimuli are presented aurally.

Difficulties in the perception of English past tense inflection were also detected by a study that used a task different from the GJT. This is presented hereafter.

Solt, Pugach, Klein, Adams, Stoyneshka and Rose (2004)

Solt et al. (2004) tested the perception of past tense regular morpheme *–ed* by L2ers to investigate whether variability in producing the morpheme could be caused by a perceptual difficulty. The authors tested 68 adult instructed learners in an immersion setting in the USA. The participants were native speakers of Mandarin, Cantonese, Russian, Spanish, Turkish, Arabic, Ukrainian or French Creole. These were divided into Low or High proficiency learners according to their scores in the Michigan Test. A special task testing the perception of regular past morpheme was designed by Solt et al. In this task, the informants listened to a teacher saying a set of sentences which involved past tense verbs and a student repeating them and they had to judge whether the repetition was the same as the original or different. In some trials, the student repeated the sentence as it was said by the teacher as in (3.3a) and in others, the student repeated the sentence without inflecting the verb for past as in (3.3b).

(3.3)

	"Teacher"	"Student"
a.	The girl *walked* in the park	The girl *walked* in the park
b.	The girl *walked* in the park	The girl *walk* in the park

The results as summarised in Table 3.9 showed that both Low and High proficiency groups had difficulty in perceiving the past tense morpheme but this was only when it occurred in coda consonant clusters. In contrast, when the morpheme occurred in non-cluster codas, it was perceived successfully more often.

Table 3.9: Perception of regular past tense inflection across three allomorphs by L2ers. (based on Solt et al., 2004, Table 2: 559)

	High proficiency group	Low proficiency group
/ɪd/ syllabic e.g., painted	90.0%	85.3%
/t/ cluster e.g., walked	70.5%	61.3%
/d/ cluster e.g., lived	62.1%	58.1%

The authors conclude that "based on these findings that L2 learners' inability to perceive the past tense *–ed* morpheme consistently across its allomorphic variants – a *systematic perceptual deficit* – is a barrier to producing this morpheme in a target-like manner" (Solt et al., 2004: 562)

What the studies by McDonald (2006), McDonald and Roussel (2010) and Solt et al. (2004) attempted to do was to explore interpretations of non-native speaker's poor performance in perception and/or processing tasks other than the lack of grammatical knowledge. Indeed, these studies managed to find evidence for the involvement of non-syntactic factors such as working memory capacity and phonological decoding ability. However, although these findings have shed light on factors that should be seriously taken into consideration when testing L2ers, whether they can explain all the observed patterns is not conclusive especially in the light of other research findings attained by more sophisticated techniques. The following study recorded ERPs during grammaticality judgment and revealed interesting results.

Chen, Shu, Liu, Zhao and Li (2007)

Chen et al. (2007) examined the receptive knowledge of subject-copula number agreement. Although it is not inflectional morphology which is tested here, this study is relevant to our investigation because we assume that the same syntactic knowledge underlies subject-verb agreement whether the verb is auxiliary or thematic. The participants of this study were 15 Chinese-speaking L2 learners of English and 15 native speakers of English who served as controls. L2 proficiency was controlled for in this study and thus the authors included only learners who had either already obtained the College English Test level 6 (the highest level) or scored higher than 80 (out of 100) on level 4 in the same test. In a GJT, participants were presented with sentences displayed visually in a word-by-word fashion on a computer screen, as in (3.4).

(3.4) a. The price of the car *was* too high. (Grammatical, congruent (G-C))
b. The price of the car*s was* too high. (Grammatical, incongruent (G-I))
c. * The price of the car*s were* too high. (Ungrammatical, congruent (U-C))
d. * The price of the car *were* too high. (Ungrammatical, incongruent (U-I))

In (3.4), while the number of the head noun (price) is kept singular in the four sentences, the number of the following copular verb (was/were) is manipulated, creating two grammatical (3.4a and 3.4b) and two ungrammatical sentences (3.4c and 3.4d). Moreover, the number of the noun (car) in the propositional phrase directly preceding the copular verb is also manipulated, creating two congruent (as in 3.4a and 3.4c) and two incongruent (as in 3.4b and 3.4d) conditions. This manipulation in number of the copular verb and the preceding noun created four sentence types as in 3.4a-through-3.4d. While ungrammatical sentences test sensitivity to violations, the (in)congruent type sentences test whether the informants' decision will be affected by the number of the nearest noun to the verb. Behavioural responses, RTs as well as ERPs, were recorded.

The accuracy results revealed that the L2 learners correctly judged sentences at a rate of 88% of the time (G-C 87%; G-I 86; U-C 90%; U-I 89%) demonstrating no statistical difference in response to different sentence types. Statistical analysis on reaction times, on the other hand, showed that the L2 learners took a significantly shorter time to respond to the ungrammatical sentences in both congruent and incongruent conditions than the grammatical sentence in the congruent condition. According to the authors, this performance represented by the high accuracy rates and faster reaction times to ungrammatical sentences demonstrates the L2 learners' ability to detect number agreement violations between the head noun and the copular verb. The ERP results, however, showed a different pattern. During the presentation of the ungrammatical sentences, the native speakers' brain waves showed LAN and P600 effects, which are taken to mean detection of the violation during automatic morphosyntactic analysis and difficulty in the integration of different types of information during reanalysis processes, respectively. The non-native participants' brain waves, on the other hand, showed neither LAN nor P600 effects in response to ungrammatical sentences, indicting insensitivity to the violation.

Chen et al.'s study reveals subtle differences between their non-native and native participants. Although the patterns arising from behavioural responses of non-native speakers resembled those of native speakers, the neural responses differed. The authors interpret these findings as an effect of the learners' native language on their SLA. They point out that "given an L1 that does not encode grammatical morphology, the learning of a syntactic agreement system in an L2 presents a major obstacle to Chinese learners of English as a second language" (Chen et al., 2007: 171).

We saw that McDonald (2006), McDonald and Roussel (2010) and Solt et al. (2004) resorted to non-syntactic factors to explain the differences between native and non-native speakers. They found evidence for the involvement of working memory capacity, lexical knowledge and access, and phonological decoding abilities. However, it seems that none of these factors can explain the results of Chen et al.'s study. This is because the methodological design of the study minimised the effects of such factors. First, the stimuli were visually presented and thus phonological decoding ability was no longer a necessary variable for completing the task. Second, only one verb (*be*) was used and manipulated in all the task stimuli, which should have immensely minimised the effect of lexical knowledge on the participants' decision. Finally, it is not clear how working memory capacity would have selectively affected neural responses but not behavioural responses represented in accuracy and RTs. It seems, therefore, that the differences captured by the ERP signatures are superior to all of these factors.

Chen et al.'s interpretation of the results as an L1 effect coincides with the interpretation of the RDH since it predicts Chinese learners to have deficits in the subject-verb agreement representations. However, based on these results alone, we cannot be sure that the observed pattern is a result of L1 effects. This is due to the fact that the non-native group consisted only of Chinese natives and we do not know if speakers from an L1 similar to English in the property tested would perform differently. The following study findings provide insights on this enquiry.

Sato and Felser (2008)

Sato and Felser (2008) investigated L2 learners' sensitivity to subject-verb agreement and case marking violations. The objective of this study was to examine to which extent the problems observed in supplying the agreement marker *–s*, as opposed to the reportedly non-problematic case marking, is reflected in sentence processing and comprehension.

The informants of this study were 20 German-speaking, 20 Japanese-speaking and 20 Chinese-speaking L2ers of English. Twenty native speakers of English were included to serve as a control group. The non-native informants' proficiency in L2 English was measured by the grammar part of the Oxford Placement Test (Allan 1992). In this test, learners scored at a range that placed them at the mid-intermediate to very advanced proficiency levels. Then, the non-native participants' knowledge of subject-verb number agreement and case marking was assessed by an offline binary-choice sentence completion task and all of them performed at a ceiling level. To test the real-time processing and comprehension of the linguistic properties under investigation, a speeded grammaticality judgement task was used. In this task, participants were presented with simple active three-word sentences on a computer screen in word-by-word fashion. They were asked to judge whether or not the stimuli sentences were well-formed and meaningful by pressing a yes or no button and their responses and reaction times were recorded. The test stimuli included ungrammatical sentences involving violations in either subject-verb agreement or case marking in addition to grammatical counterparts involving the same properties. The following are examples of the grammatical and ungrammatical sentence stimuli in agreement and case conditions.

Agreement

(3.5) a. *She rarely flirt.

b. She rarely flirts.

Case

(3.6) a. *He admires she.

b. He admires her.

In both examples, the (a) sentence creates a violation in either the subject-verb number agreement as the verb lacks the third person *–s* inflection (3.5a) or case marking as a nominative-marked pronoun replaces an objective pronoun (3.6a) and the (b) sentence is the grammatical counterpart. Sato and Felser predict that if L2 processing is influenced by a transfer of the L1 linguistic properties, the performance of the three language groups in their study might differ according to the L1-L2 similarities and differences. As German, similar to English, marks both case and agreement, the German-speaking participants are expected to perform similarly to native speakers. Since Japanese, similar to English, marks case but, different from English, lacks subject-verb agreement marking, the Japanese participants are expected to perform similarly to native speakers in case marking but to have problems with subject-verb agreement. Finally, Chinese lacks both case and agreement and therefore the Chinese-speaking informants might encounter problems with both properties

The following table (3.10) presents the accuracy rates and RTs per group in response to ungrammatical items per violation type.

Table 3.10: Mean accuracy rates % and mean reaction times in millisecond (in parentheses) per group in response to ungrammatical sentences involving violations in agreement and case (based on Sato and Felser, 2008, tables 4 and 5: 21-23)

	Agreement	Case
NS	93.8 (737)	93.1 (734)
German	93.4 (1091)	97.8 (929)
Japanese	83.8 (1292)	93.5 (994)
Chinese	62.8 (1690)	80.0 (1163)

Sato and Felser reported that statistical analyses performed on accuracy rates showed main effect for language (NSs, German, Japanese, Chinese) and violation type (agreement, case). The violation type effect was found in the mean accuracy rates of the Japanese ($p<.01$) and Chinese groups ($p<.01$) with agreement violations judged significantly less accurately than case violations. For the German learners, this effect was not found in the group mean accuracy rates ($p=.114$).

Furthermore, statistical analyses performed on mean reaction times revealed significant differences between the native speakers and each of the L2 learner groups with the former taking a significantly shorter time to detect the violation. In comparison between the reaction times divided by the violation type, the results of the three language groups (but not the NSs group) showed that they took significantly longer to detect agreement than case violations.

Sato and Felser (2008) conclude that “the learners’ previous linguistic experience does not provide a satisfactory explanation for the observed L1/L2 differences” (ibid: 26). The authors base their interpretation on the findings that all non-native speakers took a significantly longer time than native speakers to respond to stimuli involving violations and that independently of their L1 background, all non-native groups, different from native speakers, took a significantly shorter time to detect case than agreement violations. Nevertheless, the findings that all non-native speakers regardless of their L1 performed below native speakers and case violations were easier to detect than agreement violation do not mean that there is no L1 transfer. It might be true that L1 transfer cannot explain all the observed patterns but it still has an effect, as evidenced in the results presented in Table 3.10. In the agreement violations accuracy results, both Chinese and Japanese participants performed below their German counterparts (62.8%, 83.8% and 93.4% respectively). In the case of violation accuracy results, the Chinese natives performed below both the Japanese and German speakers (80.0%, 93.5% and 97.8% respectively). The RTs also seemingly show a similar symmetry. As noted above, both Chinese and Japanese lack the agreement feature and only Chinese lacks Case features. Therefore, these results do show an L1 effect; it cannot explain all the patterns in the data though.

To recap, the review in this section has shown that the difficulty with past tense and verbal agreement morphology does extend to the perception and processing modalities, but the findings are inconclusive as to whether the source of the problem is syntactic or non-syntactic. Johnson and Newport's (1989) and McDonald's (2000) studies revealed that their participants, who were native speakers of variety of L1s, were insensitive to past tense and verbal agreement violations. It was suggested that this might indicate a lack of grammatical knowledge, but the auditory presentation of the test stimuli in those two studies casts some doubt on this interpretation because the problem could be due to poor phonological decoding abilities instead. Three studies reviewed above, namely McDonald (2006), McDonald and Roussel (2010) and Solt et al. (2004), demonstrated that their informants did have phonological difficulties in the perception of the inflection. Nevertheless, insensitivity to subject-verb agreement violations by Chinese L2ers of English was attested in an ERP study conducted by Chen et al. (2007), in which the phonological factor was not involved because the test stimuli included only subject-be agreement structure. Finally, Sato and Felser (2008) investigated the role of the L1 in insensitivity to violations in verbal agreement and found that the native language could not explain their informants' performance. Despite this, however, I argued that the results did show an effect for the L1. These studies therefore have provided inconclusive evidence on the source of the perceptual difficulty of the properties under investigation leaving the issue in need of further investigation.

3.2.3 Summary and Statement of the Problem

The previous two subsections have presented and discussed previous research on the production, perception and processing (PPP) of past tense and verbal agreement morphology by adult L2ers of English. The findings were looked at in the light of different proposals on the source of morphological variability in SLA (i.e., OG, MSBH, RDH, MSIH, PTH and PhRH). I argued that to locate the source of the phenomenon under investigation, further research is needed because the data from the reviewed studies did not provide testing grounds for some of the hypotheses or the evidence advanced in those studies for or against other hypotheses was inconclusive.

The first two hypotheses in the light of which the literature was discussed were OG and the MSBH. These hypotheses propose that variability arises because of the initial absence of the underlying syntactic representations, which are built up gradually with more exposure to the target language. Furthermore, while OG asserts that the L1 does not have any effect on the building up of the L2 functional categories as they project solely based on the interaction between the linguistic input and fully available UG, the MSBH maintains that the L1 has a facilitative role in building up these projection. The literature review showed that none of the previous studies provided data suitable for testing these two proposals. This was mainly because the learners tested might have already passed the stage at which their grammars do not have the relevant projection. The lowest proficiency learners tested were Goad and White's (2006) informants, who were at the intermediate level, and as we saw they already produced the morphology consistently. To fill in the research gap in the aim of locating the source of morphological variability, these proposals need to be tested. To this end, data from low proficiency learners are required to test the proposal that variability arises because of the temporary initial absent of functional categories in SLA. Further, to test if the L1 plays a role in the building up of these categories, comparative data from learners whose native languages differ with regard to the properties under testing are needed.

The third account of the research phenomenon discussed in this chapter was the RDH. This hypothesis proposes that variability is caused by permanent absence of the syntactic representations, which is, in turn, due to a critical period effect on uninterpretable features. Hence, uninterpretable features which are not activated in FLA become permanently inaccessible in adult SLA. Evidence for or against this hypothesis came mainly from a comparison between advanced proficiency L2 speakers whose L1s contrast in having or not having the features under study. The first study testing this was Hawkins and Liszka (2003) and it provided evidence for the hypothesis but the number of participants in the study was very small, which required caution in interpreting the findings as the authors themselves pointed out. The second study was Campos Dintrans (2011). The performance of this study's informants were claimed to be evidence against the RDH as it was found that their written production was markedly higher than their oral production. However, I argued that the written production of those learners was better possibly because of their ability to control it making use of their metalinguistic knowledge. By this, we came to the conclusion that further testing for the hypothesis is needed. Therefore, to locate the source of variability, the RDH's proposal that uninterpretable feature cannot be acquired by adult learners who lack them in their native language need to be tested. This requires data from advanced proficiency L2 speakers whose L1s contrast in having or not having the features under study.

The fourth account that this chapter discussed was the MSIH. This hypothesis maintains that syntactic categories are present in SLA acquisition from early on (the source is the L1 or UG) but variability is caused by a processing failure between these categories and their surface manifestations. Evidence for this hypothesis came from data showing that variability occurs while the syntactic representations were intact. Lardiere's (2007) informant showed this pattern, supporting the MSIH. However, we observed that this proposal could not explain why L2ers from different L1 backgrounds performed differently (e.g., Hawkins and Liszka, 2003). Yet, this difference could be irrelevant to the hypothesis if the reason is found to be non-syntactic (e.g., phonological). Therefore, this hypothesis requires further testing. To locate the source of variability, the MSIH's proposal that syntactic categories are present from early on in SLA need to be tested and thus data from low proficiency learners are required.

The fifth and sixth accounts discussed in this chapter were phonological. These were the PTH and PhRH. These accounts propose that variability is promoted by L1-transferred constraints at the syllable structure (PhRH) or caused by L1-transferred constraints at the prosodic structure (PTH). First, support for the PhRH was provided by Lardiere (2007), who showed that the inflection is supplied more often in written than oral production and on regular than irregular past tense verbs, and by Campos Dintrans (2011), who showed that his informants' suppliance of the inflection was higher in consonant cluster than non-consonant cluster codas. Evidence against the PhRH came from Hawkins and Liszka (2003) and Campos Dintrans (2011), who found that their informants retained final t/d single consonants and consonant clusters in monomorphemic words more often than in regular past tense verbs. Taken together, these findings render the evidence contradictory and the issue in need of further examination. Second, evidence for the PTH came from six of Goad et al.'s (2003) informants, who produced the inflection higher in specific structures than others in line with the hypothesis. Campos Dintrans (2011) also claimed to provide evidence in support of the PTH, but as we saw it was not conclusive because it relied mainly on the difference between the written and oral production of the inflection, which could be equally attributed to factors other than prosodic constraints. These constraints need further examination as well. Therefore, to locate the source of variability, the PhRH and PTH need to be tested. This requires data from L2ers whose L1s contrast in allowing or disallowing consonant clusters word-finally. Also, since the prosodic constrains are claimed to be acquirable and thus not finding evidence in support does not necessarily mean that the hypothesis is falsified, it is preferable to test learners at different L2 proficiency levels.

Furthermore, reviewing the studies that explored the perception and processing of past tense and verbal agreement morphology in the aim of locating the source of the phenomenon, I assumed that if the source of the difficulty in the production of verb morphology is syntactic, it would extend to the perception and processing of the same morphological items. Indeed, some of reviewed studies (Johnson and Newport 1989, McDonald 2000, Sato and Felser 2008) found that L2ers suffer from difficulties in perceiving and/or processing the morphology. This finding therefore gave support to syntactic accounts of morphological variability. However, other studies (Solt et al., 2004; McDonald, 2006; McDonald and Roussel, 2010) showed that non-syntactic factors such as working memory capacity, lexical knowledge and phonological decoding abilities are at play, which was believed to cause the observed poor performance in morphology perception and processing. The latter findings provided support to non-syntactic accounts of variability. Yet, a study by Chen et al. (2007), in which the effect of these non-syntactic factors was minimised, revealed important differences in the brain signals in response to subject-verb agreement violations between native and non-native speakers, indicating that the problem could be located deeper in the syntax. Taken together, these findings render the evidence provided inconclusive and the issue in need of further investigation.

This thesis is an attempt to locate the source of morphological variability in adult SLA of English and the reason for its persistence in some learners rather than others. I aim to achieve this through testing the accounts mentioned above on the source of the problem against PPP data from Arabic or Chinese-speaking adult L2ers of English at three L2 proficiency levels (i.e., Low, Mid and High). This study design satisfies the criteria I discussed above for testing different proposals on the source of variability. Including speakers of Arabic and Mandarin Chinese allows testing syntactic and phonological proposals on the source of variability. Arabic has both syntactic features for past tense and verbal agreement and phonological structures similar to those required to accommodate the English inflection at the syllable level (i.e., coda consonant clusters), whereas Mandarin Chinese lacks both. Teasing apart these two sources of possible L1 transfer will be managed in the data collection experiments through controlling for the different phonological structures that the examined English morphemes create. The motivation for including learners at three different proficiency levels is to test the two aspects of variability which are related to its occurrence in early speech of L2ers and its persistence in the speech of some learners rather than others (see section 2.3.1 on these aspects). Finally, one important characteristic of this thesis' study is that it explores not only the production of the properties under investigation as is the case with previous research on the phenomenon but also the perception and processing of the same properties. This will provide extra light on the nature of the representations underlying the use of past tense and verbal agreement properties. The following research question will be addressed:

What is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?

It has become evident from the literature review in this and the previous chapters that the learners' native language is considered by many researchers an important affecting factor on SLA of past tense and verbal agreement. This is true at the morphosyntactic and morpho-phonological levels. Therefore, to examine the role of the L1, it is essential to describe these properties in research participants' L1s and show how they resemble or differ from the L2 properties. I set out with this in the following section.

3.3 Past Tense and Verbal Agreement in English, Arabic, and Mandarin

This section describes the morpho-syntactic and morpho-phonological properties of past tense and verbal agreement in English (the L2 target), Arabic and Mandarin (the native languages of the L2 participants) in 3.3.1, 3.3.2 and 3.3.3 respectively. In each section, the surface morphology relevant to the properties under study is presented first, followed by a description of the syntactic and phonological (at syllable and prosodic levels) structures. The review below shows that Arabic and English are similar as both mark verbs for tense and agreement and allow consonant clusters in a coda position, whereas Chinese differs from both languages as it does not mark verbs for past or agreement nor does it allow consonant clusters in any word position. It is also shown that while English prosodify the target inflections externally to the PWd violating simultaneously EXHAUST and NONREC constraints, neither Arabic nor Chinese contain structures prosodified in the same fashion. However, both Arabic and Chinese have structures that separately violate the EXHAUST and NONREC constraints.

3.3.1 Properties of English

In English, finite verbs, i.e., tensed verbs, can be specified as either [+past] or [-past]. For example, the verb *play* is [-past] in (3.7a) and [+past] in (3.7b). Past tense is expressed morphologically on regular verbs by adding the suffix *–ed* as *played* in (3.7b) and on irregular verbs by giving rise to a distinct past tense form as *went* in (3.7c). By contrast, non-finite verbs are not tensed and, thus, are not specified for [+/-past]. An example of a non-finite verb is the infinitive *see* as in (3.7d).

(3.7) a. I play the piano very well.

b. I played the piano in front of a large audience

c. I went to the concert alone.

d. I wanted to see the concert.

e. S/he sings very well.

f. I/they/we/you sing very well.

g. S/he/I/they/we/you played football.

In the present tense, English finite verbs agree in number and person with their 3rd person singular subjects by adding the suffix *–s* as in (3.7e), but the verb is used in its base form with other subjects as in (3.7f). Past tense verbs, however, do not show any agreement (3.7g).

Adopting a generative approach to language, I assume that past and agreement marking on verbs arise from a series of syntactic procedures that applies on morphosyntactic features. To illustrate these procedures, I particularly follow Adger's (2003) analysis, which is formulated within the Minimalist framework (Chomsky, 1995, 2001). Recall that in Minimalism, as elaborated in section 2.2.1, it is held that lexical items are formed from bundles of linguistic features (formal, phonological and semantic). These lexical items provide input to the computational system (syntax), which transforms them into syntactic structures by series of operations. The resulting structure interfaces with a Phonetic Form (PF) system which dictates how the structure is pronounced and a Logical Form (LF) system which gives it a meaning. Furthermore, as explained in 2.2.1, feature are referred to as interpretable or uninterpretable according to whether they play a role in the semantic interpretation or not.

Within this framework, it is assumed that assigning tense to the linguistic expression is one of the tense head (T) functions and that the appearance of tense marking on a verb (V) results from an Agree relation between T and V. This Agree relation is a feature checking and valuing procedure. Adger (2003:133-141) illustrates this as follows. The lexical verb has an unvalued uninterpretable tense feature v[*u*Infl:] (Infl stands for Inflection) and T hosts an interpretable tense feature T[past]. T Agrees with V, so when the lexical verb enters the syntactic derivation, its uninterpretable feature v[*u*Infl:] is checked and, thus, valued by the interpretable features of T[past]. Then, according to the feature Checking Requirement, once the uninterpretable feature is checked, it deletes (Adger 2003: 134). This is represented schematically as follows:

v[*u*Infl:]… T[past] ➔ *v*[*u*Infl: past]… *v*[~~*u*Infl~~: past]

Then, when the syntactic structure is spelled out (at the PF level), v[~~*u*Infl~~: past] becomes subject to a pronunciation rule that specifies the past form of the verb. Adger describes the pronunciation rule in case of regular past tense verbs as follows:

Pronounce v[~~*u*Infl:~~ past] as *ed* (Adger, 2003: 137)

However, a special pronunciation rule applies in case of irregular past tense verbs such as *eat*-*ate*. Adger describes it as follows:

> Pronounce *eat* as *ate* when it is adjacent to *v*[~~*u*Infl~~: past], and in this case, do not pronounce v[~~*u*Infl~~: past] (Adger, 2003: 137).

It can be seen here that in this analysis expressing past undergoes the same syntactic operations for regular and irregular verbs, but it differs in the pronunciation rule at the spell-out of the syntactic structure. Some researchers, e.g., Pinker and Ullman, 2002, however, maintain that different processes occur for the expression of past on regular and irregular verbs. Pinker and Ullman propose that while irregular past tense forms are listed as separate entities in the lexicon, their regular counterparts are rule driven. Verifying which analysis is on the right track is out of the scope of this thesis.

As for the present tense, according to Adger's (2003) analysis, it is computed in a similar mechanism as past tense, but with one additional procedure to ensure that the present tense verb agrees with its subject. As shown above, the present tense verb form shows different phonological forms relative to the subject of the utterance (e.g., the librarian *likes* books/ the librarian *like* books). This is what we call subject-verb agreement phenomenon. This agreement holds only in the present tense. Adger illustrates the syntactic procedures that give rise to subject-verb agreement as follows. In an utterance like '*the librarian likes books*', T hosts, among other features, an uninterpretable Phi-feature (person, number, and gender; written as *Ø*-features). This feature is checked and valued by the interpretable feature [singular] of the noun (librarian). This is schematically represented as follows

T[*uØ:*]… NP[singular] ➔ T[*uØ:*singular]… T[~~*uØ:*~~singular]

Then, when the lexical verb enters the derivation, its uninterpretable feature *v*[*u*Infl:] is checked and valued by the interpretable feature of T. This is schematically represented as follows.

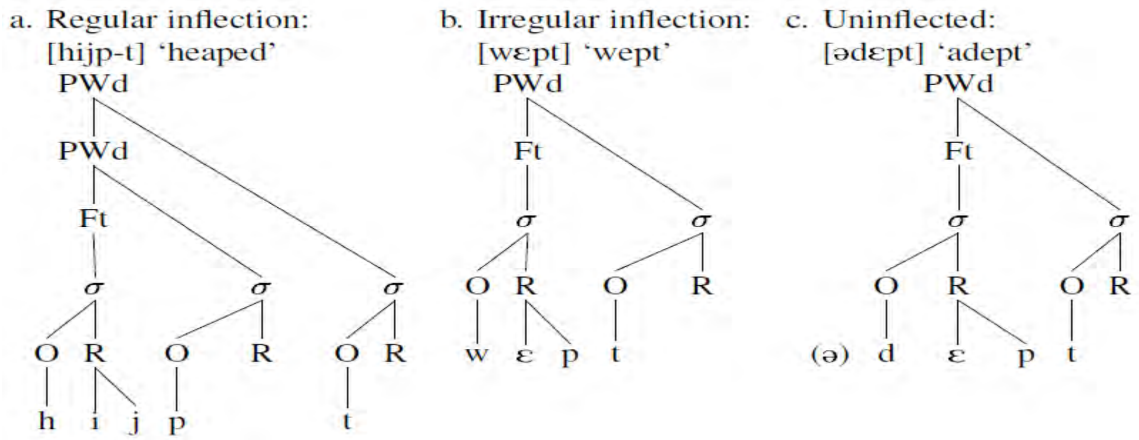
T[~~*uØ:*~~singular]… v[*u*Infl:] ➔ v[*u*Infl: singular]… v[~~*u*Infl~~: singular]

Finally, at spell-out (PF level), the v[~~*u*Infl~~: singular] becomes subject to the pronunciation rule, which ensures that the *-(e)s* is expressed on the verb (Adger, 2003: 182).

Having outlined the assumptions on the underlying syntactic structure of past and verbal agreement morphology, I turn now to describe their phonological structure at the syllable and prosodic levels. As can be observed, at the syllable level, adding the *–ed* and *–s* inflections give rise to different structures in a word coda position according to the stem shape. These are 1) VV-d/z, e.g., played /aid/, plays /aiz/, 2) CV-d/z, e.g., wanted /tid/, misses /siz/, 3) XC-t/s, e.g., talked /kt/, talks /ks/ and 4) XC-d/z, e.g., travelled /ld/, travels /lz/ (where V and C refer to vowel and consonant, respectively and X is C or V). It is obvious here that the structures in 3 and 4 create a cluster of consonants in a coda position, whereas the structures in 1 and 2 do not. These four allomorphs will be kept separate in the methodology and analyses in Chapters Five and Six in order to test the claim that the L2ers' difficulty is promoted by the nature of these structures (See PhRH section 2.3.3.6 for details on this proposal).

As for the prosodic structures of verbs inflected for past or agreement, I will adopt Goad et al.'s (2003) analysis because it bears the core assumptions of the PTH, a proposal I am testing in this thesis (see section 2.3.3.5 for full details on this proposal). In essence, the main assumption held by Goad et al. is that the regular past (*-ed*) and verbal agreement (*-s*) inflections are attached externally to the PWd violating simultaneously two constraints of the hierarchy of the prosodic structure: EXHAUST (e.g., no PWd immediately dominates a syllable (PWd-σ)), and NONREC (e.g., no PWd dominates another PWd (PWd-PWd)). By contrast, the final consonant in irregularly inflected and monomorphemic words is syllabified PWd internally, violating EXHAUST only. This is schematically represented in Figure 3.1.

Figure 3.1: Prosidification of regular inflection, irregular inflection and similar shape monomorphemic words (adopted from Goad et al., 2003: 250)



As can be seen here, 3.1a violates both EXHAUST as the PWd dominates σ and NONREC as the PWd dominates another PWd. By contrast, only the former constraint is violated in 3.1b and 3.1c. Recall that in Goad et al.'s (2003) PTH proposal, it is assumed that the structure in 3.1a is particularly difficult to L2ers whose L1s does not exhibit such a structure. Therefore, the following two sections (3.3.2 and 3.3.3) will show whether Arabic and Chinese (the L1s of the participants in this research) have a similar structure.

To sum up, past tense and subject-verb agreement are morphologically marked on English verbs. At the morphosyntactic level, adopting Adger's (2003) analysis, I assume that the lexical verb enters the syntactic derivation bearing an unspecified uninterpretable tense feature [*u*Infl:]. In virtue of the Agreement relation between V and T, the feature on V is checked and thus valued by the feature hosted in T (i.e., T[past] in past tense contexts and T[singular] in present tense contexts). At spellout, specific pronunciation rules apply giving rise to the appropriate phonological forms. At the syllable level, the addition of the *–ed* or *–s* inflections to the verb creates four different structures according to the stem shape: 1) VV-d/z, 2) CV-d/z, 3) XC-t/s, and 4) XC-d/z. At the prosodic level, following Goad et al. (2003), I assume that while the verbal agreement and regular past inflections are attached externally to the PWd violating EXHAUST and NONREC constraints, final consonants in irregular verbs and uninflected words are attached internally to the PWd, violating only the EXHAUST constraint.

In the following two sections, past and verbal agreement properties in the research participants' L1 are described. I start with Arabic hereafter.

3.3.2 Properties of Arabic

As will be detailed in Chapter four (section 4.3.1), Arab participants in this thesis are speakers of six dialects. These are Syrian, Jordanian, Iraqi, Saudi, Egyptian or Libyan. Therefore, the description in this section relates to these six dialects.

Similar to English, the verb in the six dialects appear in two distinct morphological forms in past and present contexts. Moreover, the verb agrees with its subject in person, number and gender in both past and present. This is morphologically realised as a suffix in the former and a prefix or a prefix and a suffix in the latter (see e.g., Cowell (1964) for Syrian Arabic; Al-Aqarbeh (2011) for Jordanian Arabic; Erwin (1963) for Iraqi Arabic; Al-Azraqi (1998) for Saudi Arabic; Benmamoun (2000) for Egyptian Arabic; Owens (1984) for Libyan Arabic). Table (3.11) shows the distribution of these properties on the verb *Katab* '*to write*' (prefixes and suffixes are in boldface).

Based on the data presented in Table 3.11, I will assume that Tense (past and present) and verbal agreement (Phi-features: person, number and gender) features exist in the six dialects at hand. Since my purpose here is just to verify whether past and verbal agreement features exist in the six dialects, I will not delve into explicating the syntactic procedures involved in the expression of these properties on verbs (for syntactic analyses of this, see e.g., Benmamoun, 2000 & Aoun, Benmamoun and Choueiri, 2010).

Table 3.11: The distribution of verbal agreement on past and present verb forms in six Arabic dialects

	Syrian		Jordanian		Iraqi		Saudi		Egyptian		Libyan	
	Past	Present	Past	Present	Past	Present	Past	Present	Past	Present	Past	Present
3rd p. Mas.	katab	**by**ktob	katab	**yi**ktub	kitab	**yi**ktib	katab	**yi**ktib	katab	**yi**ktib	ktab	**yi**ktəb
3rd p. Fem.	katab**ət**	**bt**ktob	kataba**t**	**ti**ktub	kitab**at**	**ti**ktib	katab**at**	**ti**kteb	katab**it**	**ti**ktib	kitb**ət**	**ti**ktəb
3rd p. Pl.	katab**u**	**by**ktob**u**	katab**u**	**yi**ktub**u**	kitb**aw**	**yi**kitb**uun**	katab**o**	**yi**kteb**uun**	katab**u**	**yi**ktib**u**	kitb**u**	**yi**kətb**u**
2nd p. Mas.	katab**t**	**bt**ktob	katab**it**	**ti**ktub	kitab**it**	**ti**ktib	katab**t**	**ti**kteb	katab**t**	**ti**ktib	ktab**ət**	**ti**ktəb
2nd p. Fem.	katab**ti**	**bt**ktob**i**	katab**ti**	**ti**ktub**i**	kitab**ti**	**ti**kitb**iin**	katab**ti**	**ti**kteb**iin**	katab**ti**	**ti**ktib**i**	ktab**ti**	**ti**kətb**i**
2nd p. Pl.	katab**tu**	**bt**ktob**u**	katab**tu**	**ti**ktub**u**	kitab**tu**	**ti**kitb**uun**	kitab**tu**	**ti**kteb**uun**	katab**tu**	**biti**ktib**u**	ktab**tu**	**ti**kətb**u**
1st p. Sing.	katab**t**	**b**ktob	katab**it**	**a**ktub	kitab**it**	**?a**ktib	ktab**t**	**?a**ktem	katab**t**	**?a**ktib	ktab**t**	**ni**ktəb
1st p. Pl.	katab**na**	**mn**ktob	katab**na**	**n**ktub	kitab**na**	**ni**ktib	ketab**na**	**ni**ktib	katab**na**	**ni**ktib	ktab**na**	**ni**kətb**u**

p= person/ Mas= masculine/ Fem= Feminine/ Pl= plural/ Sing= singular

As seen in the previous section, the target English inflections create consonant clusters word finally when added to stems of specific shapes. Furthermore, we saw that, at the prosodic level, the inflection is assumed to attach to the external PWd creating an adjunction structure. These particular structures are claimed to be difficult for L2ers whose L1s do not allow them (see sections 2.3.3.5 and 2.3.3.6 for details on these proposals). Therefore, in order to test these proposals, I set out here to demonstrate whether the six Arabic dialects allow such structures or not.

Although the syllable structure and related constraints vary from one dialect to another, some generalizations can be made in this area. The core syllable types in the six dialects are CV, CVV, CVC, CVVC, CVCC (see e.g., Cowell (1964) for Syrian Arabic; Bosisio (2003) for Jordanian Arabic; Erwin (1963) for Iraqi Arabic; Abu-Mansour (1991) & Mahfoudhi (2005) for Saudi Arabic; Mahfoudhi (2005) and Watson (2002) for Egyptian Arabic; Laradi (1983) for Libyan Arabic). While the first four types can occur in any word-position, the last type (CVCC) is restricted to a word final position. Of a particular concern to us here is this latter syllable type in which a consonant cluster is realised in a coda position. The following are examples from the six dialects showing a consonant cluster in a word-final domain (the source is provided alongside each example).

(3.8) Syrian Arabic
a. bent ‘girl’ (Cowell, 1964: 25)

(3.9) Jordanian Arabic
a. kalb ‘dog’
b. hilm ‘dream’ (Btoosh, 2006: 196)

(3.10) Iraqi Arabic
a. hind ‘India’; girl’s name
b. garb ‘west’ (Erwin,1963: 33)

(3.11) Saudi Arabic
a. galb ‘heart’
b. mahr ‘dowry’ (Mahfoudhi, 2005: 30)

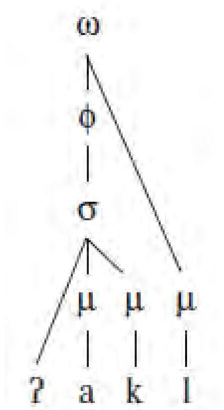
(3.12) Egyptian Arabic
a. bard ‘cold’
b. ʔalb ‘heart’ (Mahfoudhi, 2005: 30)

(3.13) Libyan Arabic
a. bint ‘girl’
b. ʒraft ‘I knew’ (Laradi, 1983: 25)

As for the prosodic structure of Arabic, there is no research (to my knowledge) on the six dialects that has demonstrated that there is a structure in which the EXHAUST and NONREC constraints are simultaneously violated (see Kiprasky (2003) for a discussion of the syllabification patterns in 15 Arabic dialects among of which the dialects of a concern here). However, there is evidence showing that there are structures that violate EXHAUST and NONREC separately under limited conditions as I shall show hereafter.

As seen above, the CV, CVV, CVC and CVVC syllable types are allowed word initially, medially or finally. By contrast, CVCC syllables are restricted to a word final position. In the light of this, it is widely agreed that the final C in such latter syllable type is extrasyllabic/ extra-prosodic, that is, it is not prosodically parsed as part of the syllable in which it occurs (see e.g., Abu-Mansour, 1991; Watson, 2002; Kiprasky, 2003; Aquil, 2013). Based on the analysis of syllabification patterns and related phenomena in 15 Arabic dialects (including the six dialects under study), Kiprasky (2003) contends that the extra-prosodic element is attached directly to internal PWd. The following proposed structure of the verb ʔakl 'to eat' in Figure 3.2 pertains to all Arabic dialects in this study.

Figure 3.2: Prosodification of the extrasyllabic C in CVCC syllables in Arabic dialects (from Kiprasky, 2003: 157, 162)

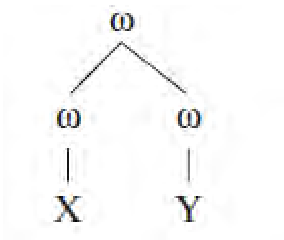


(ω = Prosodic Word; ϕ = Foot; σ = Syllable; μ = Mora)

As can be seen here the final consonant is directly dominated by the prosodic word and, hence, the EXHAUST constraint is violated. This is different from the structure proposed for English past and agreement where the inflection is attached directly to the higher PWd violating EXHAUST and NONREC simultaneously (see previous section).

Turning to NONREC, this is also violable in Arabic. The most obvious case is compound noun structures. *Raas-Maal* 'capital money' and *Abd-Allah* 'slave of God'(a name) are examples of compounds relating to the six dialects (see Ryding, 2005: 99-101 for a detailed description of compounds in modern standard Arabic, but which are still preserved in Arabic dialects). Following Ito and Mester (1992), Kiprasky (2003) posits the structure in Figure 3.3 as an analysis for compound nouns.

Figure 3.3: Prosidification of compound nouns (from Kiprasky, 2003: 154)



As can be seen here, the structure in Figure 3.3 involves a domination of a prosodic word over another, which is a violation of NONREC. Based on this, I will assume that NONREC is violable in Arabic, particularly in noun compounds. Indeed, this is the assumption held by Goad et al. (2003) for Chinese compounds as we shall see in the following section.

To sum up, it has been shown that verbs in Arabic are marked for past and present tenses and agree with their subjects in person, number and gender. Based on this, I have assumed that the six dialects instantiate syntactic features for Tense (past and present) and Agreement (person, number and gender). Then, a description of the syllable structure has demonstrated that consonant clusters are allowed in a word coda position. Finally, at the prosodic level, although there is no evidence in the six dialects for a structure that simultaneously violates the EXHAUST and NONREC constraints, structures that violate them separately do exist. On the one hand, the final C in CVCC syllables is assumed to be extraprosodic affiliating directly with the PWd and, hence, violating EXHAUST. The compound noun structure, on the other hand, is seen as a violation of the NONREC as it incurs a PWd over a PWd domination.

The Chinese properties are described next.

3.3.3 Properties of Mandarin

Different from English and Arabic, Mandarin does not mark verbs for tense or agreement and the same form of the verb appears in past, present or future contexts and with different subjects (Li and Thompson, 1989; Po-Ching and Rimmington, 2006). It should be noted that Chinese has aspect markers, which are used to denote that the action is completed (le), it is past experience (guo) or it is in progress (zai); however, "it must be stressed that aspect markers are NOT indicators of tense." (Po-Ching and Rimmington, 2006: 58; stress in original). Indeed, the verb appears in one form in all contexts, but time expressions are usually used to indicate the time reference/ context for the verb as the following example clarifies:

(3.14) a. wo zuotian jin ching qu
I yesterday into city go
'I went to town yesterday'

b. wo mingtian jin ching qu
I tomorrow into city go
'I will go to town tomorrow'

c. wo changchang jin ching qu
I often into city go
'I often went to town'

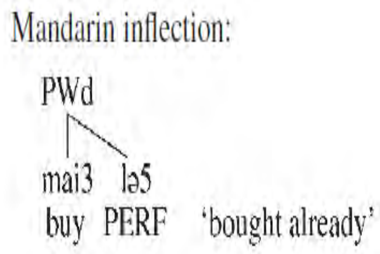
(Po-Ching and Rimmington, 2006: 74)

Based on this, I will assume that Chinese does not have syntactic features for past or verbal agreement (this assumption is also held in previous research on the topic; see section 3.2.1). This distinguishes Mandarin from English and Arabic, both of which have them.

As for consonant clusters (the structure required at the syllable level for producing past and agreement markers on English verbs), these are not only disallowed in Mandarin in a word coda position, but also "there are no consonant clusters within syllable or word margins" (Hansen, 2001: 345). This is another difference between English and Arabic on one hand and Mandarin on the other.

What remains here is to show whether Mandarin has the prosodic structure required for supplying the English inflections under study. To this end, I will adopt the analysis proposed by Goad et al. (2003) and Goad and White (2006) because again this is one of the core assumptions based on which the PTH is constructed (see section 3.2.1). Goad et al. contends that Mandarin does not allow violations of EXHAUST and NONREC simultaneously and, therefore, the adjunction to external PWd structure is not licensed in the language (ibid: 250-253). However, the authors argue that the two constraints are violable separately by different structures. While EXHAUST is assumed to be violable by the aspect marker *le,* NONREC is held to be violable by compound nouns. Goad and White (2006) adopt Duanmu's (2000) analysis of the Mandarin aspect marker as in Figure 3.4.

Figure 3.4: Prosidification of Mandarin aspect marker (adopted from Goad and White, 2006: 248)



It is clear from the structure that the aspect marker *le* links directly to the PWd in violation of EXHAUST. This is similar to the structure of final consonants in CVCC syllables in Arabic (see section 3.3.2 above). Both, however, are different from the prosodic structure of English inflections, which adjoins a higher PWd in violation of EXHAUST and NONREC.

As for the NONREC, Goad et al. maintain that this is violable in Mandarin in compound noun structures. As shown in Figure 3.3 above, the structure of compounds involves a PWd over a PWd domination violating NONREC.

In short, Mandarin Chinese does not have tense or agreement markers, which has led to assuming that it lacks syntactic representations for these properties. It has been demonstrated that Mandarin lacks the syllable and prosodic structures required for producing the English inflections under study. Consonant clusters are disallowed in a coda position and a simultaneous violation of EXHAUST and NONREC does not exist in Mandarin. The two constraints are violable separately, however, in the structures of the aspect marker and noun compounds.

3.3.4 Summary

It has been shown in this section that Arabic resembles English as it marks verbs for past and agreement, whereas Mandarin differs from both as the verb appears in its base form in all time contexts and with different subjects. This has led us to assume that English and Arabic have syntactic representations for these properties, but Mandarin does not. Then, it has been demonstrated that the addition of the English inflection to verbs of specific stem shape creates a cluster of consonants in word coda position. In this regard, it is illustrated that Arabic also is similar to English as it allows consonant clusters in word coda position. In contrast to both languages, however, Mandarin disallows such syllable structures in any word position. Finally, the prosodic structure of English inflected verbs has been assumed to involve adjunction to the external PWd violating EXHAUST and NONREC. Compared to English, Arabic and Chinese lack such prosodic structures, but in both languages EXHAUST and NONREC are violable separately. These similarities and dissimilarities are summarised in Table 3.12.

Table 3.12: Summary of syntactic features and relevant characteristics of English (L2 target), Arabic and Mandarin (native languages of research participants)

	Syntactic feature for past and verbal agreement	Consonant Clusters	Prosodic adjunction structure
English	√	√	√
Arabic	√	√	X
Mandarin	X	X	X

3.4 Order of the Following Chapters

The following chapter (four) introduces the research participants and a study conducted to measure their L2 proficiency. The next three chapters (five, six and seven) present the main studies on the production, perception and processing of past and agreement morphology. I will start with the production study in Chapter Five in accordance with most research on the topic. It will be investigated whether and at which proficiency level the Arab and Chinese participants have difficulties in producing the relevant morphological items. After that, the perception study is presented in Chapter six. Here, it will be examined if the production difficulties extend to the perception of the same participants as this provides insights into the source of the problem. Then, Chapter seven deals with the processing study; in this chapter, what will be looked at is the level of automaticity in online retrieval of the relevant properties by the research participants

Chapter 4. Measuring L2 Proficiency

4.1 Introduction

This chapter reports on how the L2 proficiency was independently measured for the participants who took part in the production, perception and processing experiments presented in the next three chapters. As shown in a previous chapter (sections 2.3.2 and 2.3.3), morphological variability has been observed to be experienced by learners at different points of L2 linguistic development and there have been different views on the role of the native language in this phenomenon. In the aim of locating the source of this target-deviant phenomenon, I decided to test L2ers of English from two different L1 backgrounds (i.e., Arabic and Mandarin Chinese) at various points of L2 linguistic development. A longitudinal design was avoided because of time and practicality concerns. A cross-sectional design was instead adopted. This design required applying a measure to gauge the participants' L2 proficiency in order to place them in relation to each other across the same language groups and in comparison to each other across the different language groups. This chapter shows how this issue was dealt with.

This chapter is organised as follows. Section 4.2 outlines the measure used to gauge the L2 proficiency of the participants in the research at hand. The participants' background and biographical information along with the task used for the collection of data relevant to the proficiency measure and the procedure followed in data transcription, coding and scoring are described in 4.3. Then, the proficiency scores are reported in 4.4. In the end of this chapter, the participants' placement into Low, Mid or High proficiency groups is given. Finally, the chapter is concluded in 4.5.

4.2 The L2 Proficiency Measure (ASCOPS)

Since the second language experience varies across L2ers, in conducting cross-group comparisons in a linguistic study there should be a means to match and compare participants with each other. In this thesis, I endeavour to test the production, perception and processing of past tense and verbal agreement functional morphology by Arabic or Chinese speakers at different points of linguistic development of L2 English and, hence, using an independent measure to place participants into proficiency-comparable groups is essential. This section outlines the proficiency measure used for this purpose and the rationale for its application in this research.

I adopted the Age-Sensitive Composite Proficiency Score (ASCOPS), which was developed by Unsworth (2005, 2008), to independently measure L2 proficiency. This measure was devised by Unsworth (2005) to measure her child and adult participants' L2 proficiency to match and compare them with each other in a study investigating the acquisition of direct object scrambling in L2 Dutch. As Unsworth (2008: 325) points out, one advantage of ASCOPS is that it "can in principle be used for any language" and therefore the application of this measure in L2 English is viable. Moreover, ASCOPS is claimed to give a proficiency score that takes age-related differences between children and adults into account. However, all of the participants in this thesis are adults. If ASCOPS can be used to measure and compare the proficiency of adults and children, there should be no doubt that it can be used with adults alone. I use it in this research for this purpose and, therefore, whether or not it is a good means for comparing children and adults is not relevant to this thesis.

The rationale for using ASCOPS rather than a standardised proficiency test (e.g. Oxford Placement Test) in this thesis is twofold. First, ASCOPS, as we will see in detail below, can be applied to measure general L2 proficiency independent of the study participants' knowledge of past tense and verbal agreement, a requirement that cannot be met in case of using standardised proficiency tests. The reason for disregarding the knowledge of past tense and verbal agreement is to avoid having a proficiency measure testing the same properties that are being tested in the experimental tasks of this research.[19] The advantage of ASCOPS over standardised proficiency tests here relates to its ability to ignore these properties altogether and obtain a final proficiency score in which the properties under investigation are not determining factors. The second rationale for using ASCOPS is that its data collection task takes less time to apply than standardised proficiency tests. As will be detailed below, a picture description task was used to collect the data for the proficiency measure; this task took only ten minutes to complete by the participants. This made the data collection session shorter and thus minimised the

[19] Unsworth (2005) excluded all utterances that had (non)-scrambled objects from the data she used to measure her research participants' proficiency in L2 Dutch because this was the same property her experimental tasks tested.

effects of participants' fatigue on the data. Therefore, for reasons of malleability in controlling the determinant factors in L2 proficiency and practicality in application, ASCOPS was held more suitable for this study than standardised proficiency tests. I turn now to describe how ASCOPS measures proficiency.

ASCOPS measures the lexical, morphological and syntactic knowledge of L2ers and combines sub-scores of the complexity and accuracy of these aspects in their oral production rendering a composite score that is assumed to indicate their general L2 proficiency in the target language (TL). Implicit in this measure is a conceptualisation of proficiency as "the ability to produce and comprehend lexically, morphologically and syntactically complex and accurate utterances in the TL" (Unsworth, 2008: 307). It is evident that the phonological aspect is neglected here; this, however, should not be a concern because this is not unusual of other proficiency tests such as cloze tests or the Oxford Placement Test, which are regularly used in the field of SLA.

In practice, ASCOPS can be used with oral production data and it gives a final score through computing and combining sub-scores from three different components. These are (i) morpho-syntactic complexity, (ii) lexical complexity and (iii) morpho-syntactic and lexical accuracy. These components are described in turn hereafter.

(i) Morphosyntactic complexity

ASCOPS adopts Ortega's (2003, cited in Unsworth, 2008: 311) definition of morphosyntactic complexity as "the range of forms that surface in language production and the degree of sophistication of such forms". Based on this, Unsworth (2008: 315) maintains that verbal density measure, which is in turn defined as "the average number of finite and non-finite verbs per T-unit [Terminable Unit]", is a good indicator of grammatical complexity. Furthermore, a T-unit is defined as "one main clause plus whatever subordinate clause and nonclausal expressions are attached to or embedded within it" (Hunt 1970, cited in Unsworth, 2008, p.314). Verbal density in the context of SLA is held to be a better measure of morphosyntactic complexity than the Mean Length of Utterance (MLU) measure, which is typically used for this purpose in L1 acquisition. This is because adult L2ers have been observed to be able to produce multi-word/morpheme utterances immediately after their initial exposure to the TL possibly due to their previous linguistic experience (Adamson 1988; Larsen-Freeman & Strom 1977, cited in Unsworth, 2008: 312). This means that MLU would be a false indicator of the grammatical complexity of such learners. By contrast, the verbal density measure is believed to reflect not only the length but also more importantly the grammatical depth of an utterance. The following sentences in (4.1) are presented to clarify this point (from Unsworth 2008: 315):

(4.1) a. She decides to go for a swim

b. The girl who is wearing a green pullover fell down

c. She shouldn't have done that

d. After the girl had eaten, she went out to play

e. The boy sits reading a book

A calculation of the finite and non-finite verbs in 4.1a-through-4.1e would reflect the grammatical complexity in an utterance such as the use of non-finite dependent clause in (4.1a), relative clause modification in (4.1b), modals, auxiliaries and complex tense/aspect forms in (4.1c and d), and durative constructions in (4.1e). In doing so, a measure of verbal density captures central aspects of grammatical development (Unsworth 2008: 315).

Therefore ASCOPS takes verbal density as a measure of morphosyntactic complexity. This is calculated as follows:

Verbal density = the total number of finite and non-finite verbs divided by the total number of T-units.

(ii) Lexical complexity

Lexical complexity is understood as the lexical diversity or richness. A traditional measure of this in both L1 and L2 acquisition is the 'Type/Token Ratio' (TTR), which is computed by dividing the number of types (vocabulary) by the number of tokens (total number of words). It is believed that a higher score of TTR indicates a richer or more diverse lexicon. For example, a learner who produces 10 Types and 20 Tokens in a speech sample is assumed to have more diverse lexicon than that of a learner who produces 5 Types and 20 Tokens and the TTR score reflects this as it would be higher for the former (10/20 = .5) than the latter (5/20 = .25) . However, Richards (1987, cited in Unsworth, 2008: 317) showed that the TTR is affected by sample size as an increase in the number of tokens would 'artificially deflate' the TTR score. For example, a learner who produces 5 Types and 20 Tokens in a sample would have a TTR of (5/20 = .25) and another learner who produces 5 Types and 30 Tokens would get a TTR of (5/30= 167). These scores inaccurately indicate that the former learner has a more diverse lexicon than the latter. Unsworth (2008: 317) explains:

> When the increase in sample size results from linguistic development within the same learner, for example when determiners are acquired, this slight dip in TTR would incorrectly suggest a lack of development (or even backsliding) whereas in reality, the learner will have made considerable steps in his or her linguistic abilities, even though the TTR does not reflect this.

Therefore, the TTR is held inadequate and 'Guiraud's index' (Guiraud 1960, cited in Unsworth 2008: 317) was adopted instead in ASCOPS. In Guiraud's index, the number of types is not divided by the number of tokens but rather by the square root of the number of tokens, which is believed to help overcoming "the problem of a negative correlation with increasing sample size (as with TTR)" (Unsworth, 2008: 317). Based on this, ASCOPS takes Guiraud's index as a measure of lexical complexity. This is calculated as follows:

Guiraud's index = number of different types (V) divided by the square root of the total number of tokens (N).

(iii) Morphosyntactic and lexical accuracy

ASCOPS also includes an accuracy measure. Unsworth (2008: 318) maintains that:

> L2ers who produce complex yet inaccurate utterances should not be considered more proficient than L2ers who produce less complex but more accurate utterances. To take the interaction between these two factors into account, a measure of proficiency should incorporate a measure of accuracy as well as a measure of complexity.

Accuracy in ASCOPS is measured by the rate of error-free utterances in a speech sample. This is calculated as follows:

Rate of Error-free Utterance = the number of error-free T-units divided by the total number of T-units in a sample.

One procedure implemented in ASCOPS before combining the results of these three sub-tests to give an overall proficiency score is standardising the raw scores of morphosyntactic complexity and lexical complexity measures. This is basically done to evaluate the participants in relation to each other in the same group. Standardised scores (or z-scores) allow comparing two or more scores positioned on a scale of the distribution of all scores from the same sample. Z-scores have a normal distribution in which 0 is the mid-point and negative and positive values are below and above average, respectively.[20]

Finally, the scores of morphosyntactic complexity (z-scores), lexical complexity (z-scores) and morphosyntactic and lexical accuracy (percentages) independent measures are computed into a single proficiency score using the 'Principal Components Analysis.' Unsworth (2005: 207) explains how this is done as follows:

> Principal components analysis is a means of reducing the number of variables in a data set, and of detecting structure between these variables. This is achieved by modeling the data on a three-dimensional scatterplot to obtain one or more new variables (or components) which account for as much variance amongst the original variables as possible. It is assumed that the original variables correlate, that is, that they measure the same construct. In our case, it is assumed that they all measure the some aspect of L2 proficiency. These variables are not identical, however, because if they were, two would be redundant. Principal component analysis is thus a means of extracting the commonalities between several variables in such a way that as much variance as possible amongst these variables is accounted for by the resulting components or factors.

The resultant scores from the principal components analysis are 'z' values that have a normal distribution, spreading on a scale where 0 is the midpoint.

To measure proficiency via ASCOPS, I collected production data through a picture description task from the participants in this research. The method is detailed in the following section (4.3) and the results are presented after that in 4.4.

4.3 Method

Following Unsworth (2005, 2008), I performed ASCOPS on semi-spontaneous production data collected from the research participants using a picture description task. This section introduces the participants in 4.3.1 and then the materials used in the picture description task in 4.3.2. After that, the procedure followed when administering

[20] Unsworth (2005) converted the raw scores from the morphosyntactic and lexical measures into standardised scores for children and adults separately. This is the procedure claimed to eliminate the age factor effect on the proficiency scores (see Unsworth (2005) for more details on this claim).

the task is presented in 4.3.3. Finally, 4.3.4 shows how the data are transcribed, coded and scored.

4.3.1 Participants

The participants of this research consist of native and non-native speakers of English. While the non-native speakers took part in the L2 proficiency measure task presented in this chapter and production, perception and processing experiments reported in the following three chapters, the native speakers participated only in the perception and processing experiment. Although the native speakers did not participate in the proficiency measure task, I introduce them in this section along with their non-native counterparts for the sake of convenience.

Informants to represent different levels of L2 proficiency were recruited as volunteers from the pool of students at a community college and a university in the north east of England by personal contact or by an email advertisement. From a total of seventy-one eventual non-native participants, nine were college students (5 Arab and 4 Chinese). Those were either partners of postgraduate students or immigrants who were following English courses for 1 to 2 years and had had English classes in their countries for 5 to 10 years. The remaining recruits were university students, four undergraduates (1 Arab and 3 Chinese) and 58 postgraduates (28 Arab and Chinese 30) enrolled in MA or PhD programs at the University of Newcastle; they reported to have had English classes for 6 to 13 years in their countries. The reason for including learners at different education levels (college vs university) was to create a wider distance between the proficiency levels at which the informants would be placed. However, this was controlled by a tendency to keep the sample as heterogeneous as possible through not including learners who significantly differed with regard to their educational background; so, all college students reported to have finished at least high-school education in their countries before arriving to the UK.

The number of non-native participants was 71 in total. A group of 10 native speakers of British English was included also to serve as controls. The non-native participants were 34 speakers of L1 Arabic (12 Syrian, 8 Libyan, 5 Iraqi, 4 Egyptian, 3 Jordanian and 2 Saudi) and 37 speakers of L1 Mandarin Chinese. The L1 background was an important criterion for the selection of the participants in this research. As explained before, this was in the aim of testing L1 effects on the research phenomenon in SLA. To achieve this, we had to choose learners from L1s that either differ from or resemble the L2 with regard to the properties under scrutiny. As elaborated in Chapter three, while Arabic is similar to English as it has the syntactic features for past and verbal agreement and it allows consonant clusters in coda position, Chinese is different from English as it is assumed to lack these syntactic features and it disallows consonant clusters. Therefore, the inclusion of such participants allows for examining the role of the native language on the acquisition of the relevant L2 properties.

Furthermore, I aimed to test learners at different points of linguistic development, so I had to measure their L2 proficiency to place them in relation and in comparison to each others. As we will see in details in section 4.4, ASCOPS placed participants at Low, Mid or High proficiency levels, which resulted in three Arab and three Chinese comparative groups. Table 4.1 presents background information related to age, years of English instruction and length of residence in the UK for each of these groups. The native speakers group is included in the table.

Table 4.1: Biographical information of participant groups

Participant groups	Age			Years of English Instruction			Length of residence in the UK		
	Range	Mean	SD	Range	mean	SD	Range	Mean	SD
Arab Low (n=11)	27-37	30	2.9	6-12	9;4	2.4	0;2-3;0	1;5	1.3
Chinese Low (n=13)	24-31	27	1.8	7-12	9.1	1.3	0;8-2;6	1;1	0.6
Arab Mid (n=14)	26-33	30	2.1	6-13	10;3	2.3	0;10-6;0	2;2	1.3
Chinese Mid (n=13)	25-36	29	3.2	9-13	10;7	1.2	0;3-8;0	1;8	2.2
Arab High (n=9)	27-34	30	2.8	6-12	9;5	1.8	0;6-8;0	2;0	2.3
Chinese High (n=11)	24-31	27;1	2	7-13	10;9	1.9	0;8-5;0	1;9	1.2
Native Speakers (n=10)	18-26	20.7	2.4	*NA	NA	NA	NA	NA	NA

** NA = Not Applicable*

The age of onset (AO) for learning English was one of the important selection criteria of non-native participants of this study. This was for the reason of reported initial age exposure effects on SLA and to test the proposal claiming that acquiring L2 uninterpretable features not instantiated in the L1 becomes impossible after puberty (i.e., RDH; see 2.3.3.3 on this proposal). Given that puberty might begin anytime between the age of 10 and 18, the safest age to take as a point after puberty is 18. This brings to light the issue of what type of input should be considered determiner of the AO. Is it the input received in English language classrooms during years of formal schooling, formal English language programmes or naturalistic exposure in the target country? In a study specifically designed to test age effects in SLA of Spanish, Granena and Long (2013: 319) set the AO for their participants as "the beginning of a serious and sustained process of language acquisition as the result of migration or the commencement of a formal Spanish language program". In this thesis, I shall adopt Granena and Long's criterion for AO. Along these terms, 'serious and sustained' exposure did not take place before the age of 18 for any of our participants. For most, this happened after they finished the high-school education and started formal English language programmes in their countries or in the UK for the purpose of achieving a level of English that would allow them to be admitted to undergraduate or postgraduate programmes. Some participants started their formal English programmes after they had settled in the UK so their naturalistic exposure to English in the UK was a better determiner of AO, but none of the participants reported to have arrived to the UK before the age of 21. Therefore, taking English formal programmes or arrival to the UK as AO satisfies the 'after puberty' criterion. Excluding English classroom exposure from the AO criterion, however, might be questioned here; our reason to do this is that no serious and sustained exposure can be considered to have taken place during these years. For all participants, classroom instruction was delivered by non-native speakers of English and their native language was used as the language of instruction (e.g., 2 hours a week × 32 weeks per academic year). All of the participants took English classes at school with the earliest starting at the age of 7 and the latest starting at the age of 15 (mean= 10;9, SD:2;0).

Finally, to ensure that the L1 factor was well controlled for, the participants were asked to give information about any linguistic experience(s) they had before other than their L1 (Arabic or Chinese) and L2 English. Only a few reported that they were beginners in a third language (French or German) at the time of the study.

Each non-native participant completed three tasks: a picture description task measuring their L2 proficiency (ASCOPS), an oral elicited imitation task testing their production of past tense and verbal agreement morphology and a computerised picture-choice task testing their perception and processing of the same morphological items.[21] The native speakers carried out only the last task.

The following section introduces the picture description task used to collect data for the proficiency measure from non-native participants.

4.3.2 The Picture Description Task

To measure the L2 proficiency of the non-native participants in this research, following Unsworth (2005, 2008), oral semi-spontaneous production data were collected using a picture description task. The materials used in this task (similar to the ones used in Unsworth (2005) but not the same), were a set of 25 pictures taken from a children's story book called *Totally Uncool* by Janice Levy (1999), illustrated by Chris Monroe. The pictures display three characters, a young girl, her father and his girlfriend. In the story book, written captions are accompanied with each picture, but for the sake of this task, written prompts were cut off. The story revolves around a young girl describing things she likes and things she does not like about her father's new girlfriend. The pictures depict general actions that the characters are involved in, such as eating ice-cream, playing music, building a sand castle on the beach and making a puppet show.

4.3.3 Procedure

The researcher interviewed the informants for ten minutes on a one-to-one basis during of which pictures were presented on a computer screen and Microsoft Office PowerPoint software was used to enable participants to go over the pictures, using the arrow keys, without the help of the researcher. Only the participants were able to view the pictures as the researcher was sitting in a chair placed opposite to them and therefore did not know which picture was going to be described, hence increasing the genuine communicative demands of the task, and in so doing minimising participants' reliance

[21] All tasks took 90 minutes in total, including two, ten-minute breaks.

on phrases and incomplete sentences to describe pictures. At first, the three characters in the pictures were introduced to participants and the main character, the young girl, was given a name, Jenny. Participants were asked to describe the pictures, or make up a story since the pictures depicted a series of related events. They were prompted to speak as much as possible by the researcher asking them questions such as “what’s next?” and giving them encouraging words such as “excellent” but this was kept to a minimum as the researcher tried not to intervene in order to avoid the participants’ production of incomplete interrupted sentences. The task was recorded using a Yamaha POCKETRAK recorder with a built-in microphone.

4.3.4 Data Transcription, Coding and Scoring

The semi-spontaneous data obtained based on the picture description task consisted of 71 sound files. The length of these files ranged between 7 and 13 minutes. To reduce variability in the amount of data collected from participants, when a file length exceeded 10 minutes, only the first 10 minutes were orthographically transcribed.

All transcriptions were made by the researcher. Following Unsworth (2005), when immediate imitations and self-corrections occurred, they were considered what the participants intended to say and, thus, were included in the data rather than the original utterances/forms. Unintelligible words[22] were coded as ‘???’, as in (4.2). T-units which had errors were highlighted and the type of error was written between brackets beside the utterance, as in (4.3).

(4.2) And then she think she need to find something ???.

(4.3) And they having some fun in the beach. [Missing *be*]

To perform the ASCOPS analysis outlined in the previous section (4.2), all T-units, error-free T-units, all verbs, different vocabulary items were counted. We hereafter look at how the contexts of each of these were counted.

[22] These were very rare.

Total number of T-units: The T-units were specified and numbered during data transcription. Types of utterances which were excluded from the counting were as follows: (1) Utterances consisting of only a proper name or a single noun, as in (4.4), (2) utterances consisting of only a prepositional phrase and containing neither a subject nor a predicate, as in (4.5), (3) one-word utterances consisting of a verb describing the action in a picture, as in (4.6) and (4) formulaic utterances, as in (4.7).

(4.4) a. Jenny.
b. The father.

(4.5) a. In the beach.
b. In the kitchen.

(4.6) a. Eating.
b. Singing

(4.7) a. Don’t know.
b. What’s that?
c. That’s it.

Error–free T-units: utterances were considered error-free unless they contained any morphological, syntactic or lexical errors. Morphological errors[23] included aspect marking as in (4.8), plural marking as in (4.9), gender as in (4.10), possessive –s as in (4.11) and case marking as in (4.12).

(4.8) a. SUB: She’s almost finish. [aspect]
Target: She’s almost finished.

b. SUB: She is talk to them. [aspect]
Target: She is talking to them.

[23] Past tense and verbal agreement morphology errors are ignored here for all participants. This is, as noted in 4.2, to avoid having a proficiency measure testing the same properties being tested in the experimental tasks reported in Chapters five, six and seven.

(4.9) SUB: The three family member are eating dinner together. [plural]
Target: The three family members are eating dinner together.

(4.10) SUB: Jenny and his father's girlfriend are sitting on the sofa. [gender]
Target: Jenny and her father's girlfriend are sitting on the sofa.

(4.11) SUB: Jenny father didn't take care of her. [possessive –S]
Target: Jenny's father didn't take care of her.

(4.12) SUB: What's her doing? [case marking]
Target: What's she doing?

Syntactic errors included non-target-like word order as in (4.13) and missing functional elements such as determiners and copula/ auxiliary *be* as in (4.14).

(4.13) a. SUB: I don't know what's this called. [word order]
Target: I don't know what this is called.

b. SUB: Why she is unhappy? [word order]
Target: Why is she unhappy? [word order]

(4.14) a. SUB: It's nice day. [missing article]
Target: It's a nice day.

b. SUB: This little girl playing games. [missing *be*]
Target: This little girl is playing games.

Lexical errors included the use of non-target-like adjectives as in (4.15), non-target-like adverbs as in (4.16), non-target-like nouns as in (4.17), non-target-like verbs as in (4.18) and non-target-like prepositions as in (4.19).

(4.15) SUB: The girlfriend looks very fashion. [adjective error]
Target: The girlfriend looks very fashionable.

(4.16) SUB: And Jenny's father new girlfriend is reading peaceful. [adverb error]
Target: And Jenny's father new girlfriend is reading peacefully.

(4.17) SUB: He has a scorn of ice-cream. [noun error]
Target: He has a horn of ice-cream.

(4.18) SUB: The woman who put on umbrella.[24] [verb error]
Target: The woman who holds an umbrella.

(4.19) SUB: And she looks for him. [preposition error]
Target: And she looks at him.

All Verbs: All finite and non-finite verbs were counted. This counting included copula and auxiliary *be*, and *do* and modals in addition to all lexical verbs. Non-target-like verbs, such as the verb exemplified in (4.18) above, were also included. Following Unsworth (2005), the rationale for this was to avoid penalising the same error twice; utterances that contained non-target-like verbs were marked as inaccurate, and if these verbs were also excluded from counting, the participant would have been penalised twice for using a verb in a non-target-like way.

Different Vocabulary Items: This counting included only vocabulary items upon their first occurrence. This was done by the text statistics function in the NoteTab software (Fookes Software Ltd).

Finally, after counting all of these contexts, the statistical operations mentioned in 4.2 were performed. Detailed results of the three sub-scores and the overall proficiency scores are reported in the following section.

4.4 Results: Proficiency Scores

As elaborated in section 4.2, ASCOPS comprises morphosyntactic complexity, lexical complexity and morphosyntactic and lexical accuracy measures and it combines the scores of these three independent measures to give an overall proficiency score. This

[24] This utterance noticeably contains other errors.

section presents the results of the three measures, the overall proficiency scores and how participants were divided into Low, Mid and High proficiency groups.

4.4.1 Morphosyntactic Complexity

As elaborated in 4.2, verbal density in the speech of participants was taken as an indicator of morphosyntactic complexity. This was measured by calculating the total number of finite and non-finite verbs and dividing it by the total number of T-units in the speech sample collected from each participant. This gives us the average number of verbs in a T-unit.

Our participants' scores on this measure were as follows. The verbal density scores ranged from 1.25 to 2.57 verbs per T-unit for the Arab participants (mean = 1.95; SD = 0.33) and from 1.47 to 2.62 verbs per T-unit for the Chinese participants (mean = 1.91; SD = 0.28). These results were then converted into standardised (z) scores (the rationale for this is explained in section 4.2).

4.4.2 Lexical Complexity

As elaborated in 4.2, lexical complexity is measured by 'Guiraud's index', which is a Type/Token Ratio variant. This is calculated by dividing the number of different vocabulary items by the square root of the total number of words of the speech sample collected from each participant. This gives us a lexical diversity score for each participant.

Our participants' scores on this measure were as follows. The lexical diversity scores ranged from 5.30 to 9.20 for the Arab participants (mean = 7.25; SD = 0.76) and from 6.63 to 9.13 for the Chinese participants (mean = 7.71; SD = 0.63). These results were then converted into standardised (z) scores (the rationale for this is explained in section 4.2).

4.4.3 *Morphosyntactic and Lexical Accuracy*

As elaborated in 4.2, accuracy in the speech of L2 learners was measured by the rate of error-free utterances. This is calculated by dividing the number of error-free T-units by the total number of T-units in the speech sample collected from each participant.

Our participants' scores on this measure were as follows. The rate of errorfree utterances ranged from 47.1% to 100% for the Arab participants (mean = 76.8%; SD = 11.9%) and from 46% to 94.1% for the Chinese participants (mean = 77%; SD = 10.1%).

4.4.4 *Computing a Single Score and Grouping Participants*

As shown above, two of the sub-scores are standardised (z) scores and the third one is a percentage. These were combined into a single score by a statistical means called 'principal component analysis' used by Unsworth ad described in section 4.2. The resulting scores were standardised (z) values, which were used as the overall proficiency scores for each participant. The scores ranged from -2.60 to 2.03. These scores were used to group participants in different-L1 comparable groups and similar-L1 different proficiency groups as shown hereafter.

The overall proficiency scores are 'z' values that have a normal distribution, spreading on a scale where 0 is the average score, positive values are above average and negative values are below average. Following Unsworth (2005), participants who score between -.50 and .50 are grouped as Mid-level learners and participants who score below -.50 and above .50 are grouped as Low-level and High-level learners respectively. This division produces three Arab groups and three Chinese groups matching in proficiency. The following table presents the proficiency data for the Arab and Chinese groups with Mann-Whitney test scores obtained upon comparing the Arab and Chinese groups that are assumed to match in L2 proficiency.

Table 4.2: Overview of proficiency groups

	Arab			Chinese			Mann-Whitney tests comparing Arab/Chinese groups
	N	Mean	SD	N	Mean	SD	
Low	11	-1.08	62	13	-0.97	38	Z= -.261, p=.794
Mid	14	0.02	33	13	-0.06	28	Z= -.825, p=.409
High	9	1.20	60	11	1.29	49	Z= -.798, p=.425

As can be seen here, the proficiency scores' means of Arab and Chinese similar-proficiency groups are close to each other and the Mann-Whitney tests reveal no significant differences between them. The lack of statistical difference therefore provides a reliable basis for further analysis between different L1 groups, based on assumption of homogeneity of proficiency within each level (for individual proficiency scores, see appendix D).

4.5 Conclusion

In this chapter, the study conducted to measure the L2 proficiency of the participants of this research and its results have been reported. It has been demonstrated that ASCOPS (Unsworth, 2005, 2008), the L2 proficiency measure adopted in this thesis, combines sub-scores of morphosyntactic and lexical complexity and accuracy in order to give a general L2 proficiency score. The rationale for using it in the present research has been explained as that the test allows excluding the properties under study in the main experiments (past tense and verbal agreement) to be determining factors in L2 proficiency. This was in order to avoid having a proficiency measure testing the same knowledge that is being tested in the main experiment. The research participants and the picture description task used to collect semi-spontaneous oral data from them have been described. Then, the proficiency scores of research participants, resulting from applying ASCOPS on the collected data, have been presented. Finally, it has been shown how the participants were placed in comparable proficiency groups at Low Mid and High levels.

In the following three chapters, we will see how these proficiency scores are put to use in the analysis of the L2 production, perception and processing of past tense and verbal agreement.

Chapter 5. The Production of Past Tense and Verbal Agreement: The Study Design and Results

5.1 Introduction

This chapter reports on the study conducted on the production of past tense and verbal agreement. As is shown in Chapters two and three (see sections 2.3.2 & 3.2), L2ers of different languages, including French, German and English, are widely observed to produce the verbal tense and agreement morphology optionally. One of the main characteristics of this phenomenon that has particularly been highlighted is that not only is it experienced by L2ers at early to intermediate proficiency levels (e.g., Prévost and White, 2000a, b) but also it sometimes persists into advanced stages of proficiency by speakers of some L1 backgrounds (e.g., Hawkins and Liszka, 2003). The current study examines the phenomenon in L2ers of English at Low, Mid and High proficiency levels.

As shown in Chapter two (sections 2.3.2 and 2.3.3), while there is no disagreement on the occurrence of a stage/s of morphological variability in SLA, the interpretation given to it is hotly debated. Some researchers attribute the target-deviant phenomenon to temporary (Vainikka and Young-Scholten, 1994, 2011, 2013; Hawkins, 2001) or permanent absence of syntactic representations (Hawkins and Liszka, 2003), whereas others propose non-syntactic accounts such as a processing failure in production (Prévost and White, 2000a, b) or phonological difficulties (Lardiere 1998a, b; Goad et al. 2003; Goad and White, 2006). The present chapter and following two chapters examine the phenomenon in the aim of resolving these debates and locating the source of variability and its persistence. To this end, an experiment testing the production of past and verbal agreement was designed. An elicited imitation sentence task was used to collect data from Arabic and Chinese learners of English who were matched in L2 proficiency at Low, Mid and High levels.

The structure of this chapter is as follows. Section 5.2 presents the experimental hypotheses, which will be formulated based on the accounts of morphological variability introduced in Chapter two (section 2.3.3). Section 5.3 outlines the methodology of the experiment designed to test the production of past tense and verbal agreement. The results are presented in section 5.4. Finally, the chapter is concluded in

5.5 with summary of the findings and evaluation of the experimental hypotheses in the light of the reported results.

5.2 Research Hypotheses

The main research question addressed in this thesis is presented in Chapter one and repeated below for convenience.

> *What is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?*

To answer this question, six experimental hypotheses have been formulated. The formulation of these hypotheses is based on a number of considerations relating to the L2 proposals presented in section 2.3.3, the literature on past tense and verbal agreement discussed in section 3.2, the similarities and differences between the L1 and L2 of the research participants presented in section 3.3, and the L2 proficiency levels of the research participants as measured in Chapter four.

First, Organic Grammar (OG; Vainikka and Young-Scholten, 1994, 1996, 1998, 2011, 2013; see 2.3.3.1) and the Modulated Structure Building Hypothesis (MSBH; Hawkins, 2001; see 2.3.3.2) similarly propose that the source of morphological variability is the temporary absence of the syntactic representations for past and agreement, which are built up gradually with more exposure to the target language and based on the interaction between the linguistic input and fully accessible UG. These two proposals however differ in their views on the role of the native language in the building up of the relevant representations. While OG maintains that no L1-to-L2 transfer takes place at the domain of functional categories, the MSBH argues that the presence of the representations under scrutiny in the L1 precipitates their acquisition in the L2. The participants in this research are native speakers of Arabic or Chinese. As shown in Chapter three, Arabic is assumed to have the syntactic representations for past tense and verbal agreement, whereas Chinese lacks them. Therefore, for OG and MSBH, syntactic representations for the properties under scrutiny are assumed to be absent from the underlying grammars of the research participants at lower proficiency levels and be built up in higher proficiency level participants. The morphosyntactic similarities and differences between the L1s and L2 of the research participants are expected not to generate differences between the different L1 participants according to OG but give advantage to Arabic participants over their Chinese counterparts according to the MSBH. This leads to the formulation of hypotheses one and two based on OG and MSBH, respectively, as follows:

H1: *The production of morphology will be variable at lower proficiency levels in Arab and Chinese participants alike, and both language groups will improve similarly with rising proficiency.*

H2: *The production of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, but Arab participants will improve faster than their Chinese counterparts with rising proficiency.*

We have seen in Chapter two (2.3.3.3) that the Representational Deficit Hypothesis (RDH; Hawkins and Liszka, 2003) also attributes morphological variability to the absence of the syntactic representations. It specifically proposes that uninterpretable features that are absent from the L1 cannot be acquired in adult SLA. Chapter three (section 3.3) showed that the properties under study are associated with uninterpretable features and while Arabic, similar to English, has them, Chinese does not. Therefore, these syntactic features are available for the Arab participants but unavailable for and cannot be acquired by the Chinese participants according to this hypothesis. This leads to the formulation of hypothesis three as follows:

H3: *Arab participants will have an L1-based advantage over their Chinese counterparts in the production of morphology from lower levels onwards.*

The fourth proposal discussed in Chapter two (2.3.3.4) is the Missing Surface Inflection Hypothesis (MSIH; Prévost and White, 1999, 2000). This is a full transfer/ full access account, that is, both full transfer of the L1 linguistic system and full access to UG are assumed to take place in SLA. Morphological variability is thought to be caused by a processing failure between the fully present representations and their surface realizations due to communication pressure during L2 production. Therefore, syntactic representations for the properties under scrutiny are assumed to be available to different-L1-speaking L2ers alike (through L1 or UG). This implies that similarities or differences between the L1 and L2 of learners in the domain of morphosyntax are not expected to generate difference in L2 performance of learners from different L1 backgrounds. However, this does not preclude that differences between the L1 and L2 in domains other than morphosyntax such as phonology cause difficulties in SLA and thus generate differences between learners from different L1 backgrounds. This leads to the formulation of hypothesis four as follows:

H4: *The production of morphology by Arab and Chinese participants at the same L2 proficiency level will be similar.*

Two phonological hypotheses were introduced in Chapter two. The first one was the Prosodic Transfer Hypothesis (PTH; Goad et al., 2003; Goad and White, 2006). As elaborated in section 2.3.3.5, the PTH attributes morphological variability to L1 transferred constraints on the prosodic structure of the L2 inflection. This takes place when the L1 does not allow the prosodic structure required for prosodifying the target inflection. Goad et al., (2003) and Goad and White (2006) propose that the English past (*–ed*) and verbal agreement (*–s*) inflections are prosodified PWd externally creating a prosodic adjunction structure that violates EXHAUST and NONREC constraints simultaneously. However, irregular past tense verbs involve prosodification to the PWd internally violating the former constraint only. The authors hypothesise that L2ers of English would experience difficulty with producing these inflections if their L1s does not allow the adjunction structure. As shown in section 3.3, neither Arabic nor Mandarin Chinese (the research participants' L1s) allows this structure, but both allow prosodification to PWd internally. Moreover, although EXHAUST and NONREC constraints cannot be simultaneously violated in either language, they are violated separately in different structures. Therefore the two language groups in this research are expected to have the same difficulty with the English inflections. Yet, according to the PTH, they should not experience difficulty with the inflection when it can be prosodified PWd internally. In addition to irregular past tense verbs, according to Goad et al., the PWd internal structure is possible under three other conditions. These are presented in Chapter three (section 3.2.1) and repeated here for ease of reference:

1- Inflection as onset: the inflection is followed by a word that starts with a vowel so the inflection can be syllabified as an onset as in "builds on" [bIldzan].

2- Inflection as coda: the inflection is added to a base ending with a sibilant so a

> schwa epenthesis occurs and it can by syllabified as a coda as in “races” [rejsəz].
>
> 3- Inflection as foot-internal: the inflection is added to a base which is ….VX] in shape (X = V or C) so it can be prosodified to the foot as in “fills” [fIlz] or “sews” [sowz].
>
> (Goad et al., 2003: 257)

Accordingly, this leads to the formulation of Hypothesis five as follows:

> H5: *The production of morphology by Arab and Chinese participants at the same L2 proficiency level will be similar and the production of morphology will be higher on verbs where the inflection can be prosodified word-internally than on verbs where it cannot.*

The second phonological hypothesis introduced in Chapter two is the Phonological Reduction Hypothesis (PhRH; Lardiere 1998a; 2003). As shown in 2.3.3.6, the PhRH proposes that phonological constraints on the syllable structure, specifically on consonant clusters, exacerbate morphological variability in the use of past and verbal agreement morphology. These constraints are also thought to be imposed by the learners’ L1. As described in section 3.3.1,The English *–ed* and *–s* morphemes occur in four allomorphic variants, two of which create a consonant cluster structure in a word final position (e.g., talked /kt/, talks /ks/ & traveled /ld/, travels /lz/) and the other two do not (e.g., played /eid/, plays /eiz/ & wanted /ted/, loses /zez/). It is the consonant cluster structure in coda position that the PhRH claims to be difficult for L2ers whose L1s disallow it. In section 3.3, it was demonstrated that Arabic and Chinese diverge in this regard; while Arabic allows consonant clusters to occur in a coda position, Chinese disallows them in any word position. This leads to the formulation of hypothesis six as follows:

> H6. *Arab participants will have an L1-based advantage over their Chinese counterparts, who will produce the morphology higher on non-cluster than consonant cluster verbs and irregular than regular verbs.*

5.3 Method

5.3.1 Design Overview

This experiment tested the production of English past tense and verbal agreement morphology by Arabic or Chinese speakers at three proficiency levels of L2 English- Low, Mid and High. The experimental task is a sentence elicited imitation task in which participants are presented with oral sentences and asked to repeat them. The test stimuli included verbs which created various phonological structures word finally when the inflection was added to them.

5.3.2 Participants

The participants in this study are 71 learners of L2 English. These are 34 native speakers of Arabic and 37 native speakers of Mandarin Chinese. Their L2 proficiency was measured by ASCOPS (Unsworth 2005, 2008) and accordingly they were placed at Low (N= 11 Arab & 13 Chinese), Mid (N= 14 Arab &13 Chinese) and High (N= 9 Arab & 11 Chinese) proficiency levels. These participants' background and L2 proficiency information are presented in Chapter four. Therefore, I refer the reader to Chapter four, section 4.3.1, for a description of the participants and section 4.4 for details on their proficiency.

5.3.3 Elicited Imitation: Theoretical Assumptions and Design Considerations

The use of the elicited imitation method has been often criticised as it could result in verbatim imitation and thus the elicited data might not represent the linguistic competence of test-takers. However, researchers agree that a careful design of the task obviates the shortcomings of this method (Vinther, 2002). Before describing the task of this study and in the aim of validating it, I start hereafter with a brief discussion of the theoretical assumptions underlying the Elicited Imitation method and highlight some important design-related issues that should be taken into consideration in order to ensure that the method taps L2 competence.

In an elicited imitation (EI) test, research participants are presented with an oral stimulus which is either read aloud by the experimenter or aurally presented from a pre-recorded tape and then participants are asked to repeat what they have heard as exactly as possible. Similar to many other data gathering techniques in SLA research, the EI technique seeks to tap the linguistic competence of language learners. The main assumption underlying this process of imitating a presented stimulus is that only if the grammatical structure to be tested is part of the test taker's linguistic competence will it be imitated correctly (Mackey and Gass, 2005). Moreover, the EI technique is assumed to be reconstructive; that is, it involves processing, comprehending, and reformulating the stimulus relying on the linguistic system of the test-taker. Describing the underlying processes of perceiving and repeating a stimulus, Bley-Vroman and Chaudron (1994: 247) write: "[…] the subject hears the input and processes it, forming a representation […] the resulting representation includes information at various levels […] the representation must be kept in short term memory […] the subject formulates a sentence based on the accessed representation."

Nevertheless, a major concern among researchers is that the nature of the EI technique makes the presented stimulus subject to rote imitation; that is, test-takers parrot what they hear. In such cases, data collected using an EI technique are not reflective of anything but the test taker's ability to repeat presented stimuli verbatim (Vinther, 2002). In this thesis, it is the reconstructive EI that we seek. Then, how can we ascertain that an EI task is reconstructive and, thus, a gauge of linguistic competence? Erlam (2006: 467) answers this as she writes that "the **design** of an elicited imitation test can largely determine to what extent it is either a measure of a learner's internal language system or a measure of his/her ability to imitate given stimuli verbatim" (emphasis mine). Erlam's answer is widely agreed upon in the research literature. There are some features that can be manipulated in the design of an EI test to require test-takers to attend to the stimulus, but process it for meaning and then rely on their linguistic knowledge system to repeat it. According to Erlam (2009), for an EI test to be reconstructive three features should be carefully controlled in its design: 1- Length of stimuli, 2- focus on meaning rather form and 3- delayed imitation. I briefly elaborate on each hereafter.

1- Length of Stimuli: This feature is often manipulated in the design of an EI to prevent the participants from memorizing the exact wording of a stimulus and the likelihood of parroting it without resorting to their linguistic knowledge system. A short stimulus is believed to be easy to store and recall from short-term memory and thus the length of test stimuli should be "beyond the capacity of their [test-takers] immediate memory span" (Hamayan, Saegert and Laraudee, 1977: 86). The appropriate length of the stimuli according to Ellis and Barkhuizen (2005: 38) would require participants "to re-encode the meaning using their own linguistic resources". The assumption here is that when the stimulus length surpasses the short-term memory span, a test-taker resorts to her/his grammatical system to reformulate and produce the target utterance. Then, how could the researcher determine the appropriate length of stimuli? Ellis and Barkhuizen (2005) point out that the appropriate length depends on the proficiency of a test-taker and, accordingly, what is suitable for one participant might be unsuitable for another. This dilemma is usually resolved by including sentences of various lengths within the same test.

2- Focus on meaning rather form: Erlam (2009) proposes that for an EI test to be reconstructive and avoid rote-imitation-based performance, it should draw the test takers' attention to meaning rather than language form. Erlam's proposal was advanced based on the findings of two previous studies. The first study was Hulstijn and Hulstijn (1984), which showed that test-takers would focus on form when they are informed to do so and the second study was Murphy and Shapiro (1994), which found that the likelihood of retaining surface forms was enhanced when the "task demands" required focus on form rather than meaning (both cited in Erlam, 2009: 70-71). According to Erlam, such findings require researchers to include a certain procedure in the EI to draw their informants' attention to the meaning of test stimuli to minimise the chance for verbatim repetition.

3- Delayed imitation: This procedure is sometimes controlled in the design of an EI to impose a time interval between the presentation of the stimuli and the repetition of the same. Previous research has demonstrated that the likelihood of imitating the stimuli verbatim increases when the test requires participants to repeat the stimuli immediately, but it decreases when a short time of 3-to-5 seconds intervenes (e.g., McDade, Simpson and Lamb, 1982). However, this time interval might be exploited by participants to plan their utterances (Vinther, 2002), and this is why it is preferable to complete the test under time pressure. Yet, a time interval does not preclude the possibility of performing the test under time pressure as both can be retained together; that is, a short time interval can be imposed but simultaneously another procedure can be included to engage test-takers with something else other than planning their responses. Erlam (2009), for instance, imposed a time interval during which participants were required to give their beliefs on the test sentences as to whether they were true or not true before repeating them.

The following section introduces the EI test used in this study and shows how these features were taken into consideration in its design.

5.3.4 The Sentence Elicited Imitation Task: Materials

In the EI task of this study, participants were given 45 aural sentences accompanied by 45 3-picture sets presented on a computer screen and they were asked to choose the correct picture before repeating the sentence. How the EI task was designed and administered to informants are described below.

Test content: The EI test consisted of 45 sentences, which contained 67 verbs testing past tense and verbal agreement morphology. These verbs were chosen to represent the irregular past tense forms as well as the four different word-final phonological structures that arise when verbal agreement *–s* and regular past *–ed* inflections are added to verbs (i.e., VV-z/d, CV-z/d, C-s/t, C-z/d) (for more details on these structures see Chapter three, section 3.3.1). From the 45 sentences, 30 were grammatical and the other 15 were ungrammatical (rationale below) as they contained 15 ‘non-inflected’ verbs in past tense or verbal agreement obligatory contexts. The 67 tested verbs are shown in Table 6.1, and their contexts and forms are controlled as follows:

- Thirty-Six verbs in past tense contexts from the following forms:
  - (I) Five verbs with a ‘VV-d’ word-final phonological structure.
  - (II) Five verbs with a ‘CV-d’ word-final phonological structure.
  - (III) Five verbs with a ‘C-t’ word-final phonological structure.
  - (IV) Five verbs with a ‘C-d’ word-final phonological structure.
  - (V) Eight irregular verbs.
  - (VI) Eight non-inflected verbs in obligatory contexts (i.e., ungrammatical).

- Thirty-one verbs in third-person agreement contexts from the following forms:
  (VII) Six verbs with a ‘VV-s’ word-final phonological structure.
  (VIII) Six verbs with a ‘CV-s’ word-final phonological structure.
  (IX) Six verbs with a ‘C-s’ word-final phonological structure.
  (X) Six verbs with a ‘C-z’ word-final phonological structure.
  (XI) Seven non-inflected verbs in obligatory contexts (i.e., ungrammatical).

Table 5.1:All tested verbs in the sentence elicited imitation task

	Agreement				Past				
	VV-z	CV-z	C-s	C-z	VV-d	CV-d	C-t	C-d	Irregular
Grammatical	Buys studies goes stays goes pays	Finishes loses washes misses watches brushes	Gets meets likes hates wants speaks	Learns reads cleans owns sells feeds	Borrowed played tried cried enjoyed	Visited wanted invited painted needed	Stopped asked walked passed helped	Travelled returned arrived called lived	Bought spent ate went saw broke had gave
Ungrammatical	Play, stay, look, wake, take, seem, feel				Stay, study, talk, ask, kill, open, take, catch				

Grammatically incorrect sentences were included in the test in order to examine whether participants would spontaneously correct some of the errors as an indication that the test has been reconstructive (Erlam, 2006, 2009). The following are examples of grammatically correct (5.1) and grammatically incorrect (5.2) sentences (a list of all sentences can be found in appendix E).

Grammatical:

(5.1) a. Yesterday, Sam **borrowed** some money and **bought** a computer.
 b. Every year, Bob **buys** a new car and **sells** the old one.

Ungrammatical:

(5.2) a. Last night, the police **catch** two thieves robbing a bank.
 b. Sam **play** chess every day but he always loses.

To create an obligatory context for past tense or verbal agreement, each of the test sentences started with or contained an expression explicitly referring to past, such as *yesterday*, *last night* and *last year* or an expression clearly denoting present tense such as, *every day*, *every year* and *always*.

Sentence length and complexity: First, the test sentences varied between 8 and 16 syllables in length (Mean = 12 syllables). As elaborated in the previous section, a short stimulus is subject to rote imitation and long stimulus might be too difficult to retell. What makes a stimulus too short or too long is the participant proficiency level and thus what is suitable for one participant might not be suitable for another (Ellis and Barkhuizen, 2005). Therefore, sentences of various lengths were included here to avoid this problem in line with previous research on the method (e.g., Erlam 2006, 2009). Second, the test design did not control for complexity but the sentences incorporated a mix of simple and compound structures.

Picture choice: Focus on meaning: I aimed in designing this task to get participants to focus on meaning rather than form to ascertain that the elicited data represent participants' linguistic competence rather than their ability to imitate the stimulus verbatim. As mentioned in the previous section, researchers agree that for an EI test to be reconstructive, participants must comprehend the stimulus before repeating it. To ensure this, Erlam (2009) constructed her test sentences as statements that could be true or false and asked her participants to give their beliefs (e.g., true, not true or not sure) for each sentence before repeating it. In this study, it was not possible to follow Erlam's procedure due to the nature of the sentences as they were not statements that could be agreed or disagreed with. Therefore, a new procedure was employed as follows. The test was designed as a picture-choice task, that is, with each sentence, three pictures were shown. One of the pictures was a true depiction of the accompanying sentence and the other two were distractors. The participants were asked to choose the correct picture before repeating the stimulus. The following is an example of a sentence-picture trial.

(5.3) Yesterday, John ate his breakfast before he went out.

Stimuli administration and delayed imitation: The test was designed in such a way as to ensure consistency in the administration of the test sentences-picture trials and not to allow participants time to rehearse their utterances. Accordingly, a short time interval was imposed after the presentation of the sentence to delay imitation, and during this interval, participants were engaged in choosing one of the pictures. The experiment was designed and run using MATLAB software (The MathWorks Inc. 2008). The final presentation design of the test was then as follows: first, the three pictures appeared on a computer screen and the sound file of the test sentence started. Immediately after that a beep sound of 200 ms duration played to alert participants to choose one picture by pressing one of the computer keyboard number buttons 1, 2 or 3, which corresponded to the three presented pictures. Once a picture was chosen, another beep sound of 200 ms duration went off to alert participants that it was time for them to repeat the sentence they heard. Moving to the next trial required pressing a keyboard button.

In addition to the 45 experimental sentence-picture trials, five sentences (three grammatical and two ungrammatical) were constructed and attached to five sets of pictures to be provided to participants as training before performing the main task. All test sentences were recorded by a male native speaker of British English and then digitized. All of the visual stimuli used in this task were taken from Google images at http://www.google.co.uk/imghp.

5.3.5 Procedures

The EI was the second task to be completed by research participants after the picture description task used to measure their proficiency described in Chapter Four. First, participants were guided by the researcher through the 5 training trials. They were asked to choose the picture that they believed to be the one described by the aurally-presented sentence after they had heard the first beep sound and to repeat the sentence after they had heard the second beep sound. They were also told to repeat the sentences in correct English but they were not told at any time that they would encounter grammatically incorrect sentences. Two of the training session trials (one grammatical and one ungrammatical) were modeled by the researcher to show participants how their responses were expected to be. After the training session, participants confirmed they clearly understood what they were asked to do. The remainder of the test was divided into three batteries and each battery comprised 15 sentence-picture trials which took 4 to 5 minutes to complete. Each battery contained three 8-syllable, three 10-syllable, three 12-syllable, three 14-syllable and three 16-syllable sentences. These were presented in the order from shorter to longer. Participants' responses were recorded on a Yamaha POCKETRAK recorder with a built-in microphone.

Elicited data were scored according to two criteria:

i. Obligatory context for past tense or third-person agreement created – supplied
ii. Obligatory context for past tense or third-person agreement created – not supplied

The first criterion comprised responses where the participant created an obligatory context for past tense or third-person verbal agreement and supplied the target form of the verb. An obligatory context was defined as a context where native speakers could use no form of the verb other than the inflected verb for past tense or verbal agreement. The second criterion comprised responses where the participant created an obligatory context for past tense or verbal agreement but used the bare form of the verb (i.e., non-target).

Two further coding procedures were performed on the data in accordance with the Phonological Reduction Hypothesis (PhRH; Lardiere, 1998a, 2003) and the Prosodic Transfer Hypothesis (PTH; Goad et al. 2003; Goad and White, 2006) and in the aim of testing Hypotheses five and six formulated in section 5.2. These were related to the syllable and prosodic structure of the verb. First, as for the syllable structure coding, verbs were grouped according to the shape of the final syllable after adding the past tense or verbal agreement inflections. These were as follows for verbs in past tense contexts: (1) VV-d (e.g., played /eid/), (2) CV-d (e.g., wanted /ted/), (3) C-t (e.g., asked /kt/) and (4) C-d (e.g., travelled /ld/). For the verbal agreement, the coding was as follows: (i) VV-z (e.g., plays /eiz/), (ii) CV-z (e.g., loses /zez/), (iii) C-s (e.g., meets, /ts/) and (iv) C-z (e.g., reads /dz/). Second, as for the prosodic structure coding, verbs were grouped according to the prosodic analysis of the past tense or verbal agreement inflection as follows: (A) No option for the inflection to be prosodified word-internally and (B) the inflection can be prosodified word-internally. Following Goad et al. (2003: 257), the inflection was deemed possible to be prosodified word-internally if it occurred in one of the following contexts: (1) the inflection is followed by a word that starts with a vowel so the inflection can be syllabified as an onset as in “builds on” [bɪldzan], (2) the inflection is added to a stem ending with a sibilant so schwa epenthesis occurs and it can by syllabified as a coda as in “races” [rejsəz] and (3) The inflection is added to a stem which is ….VX] in shape (X = V or C) so it can be prosodified to the foot as in “fills” [fɪlz] or “sews” [sowz].

The coding was also applied to the grammatically incorrect sentences that were included in the test. Recall that the design of grammatically incorrect sentences was similar to that of the correct ones and the only difference was that verbs in the incorrect sentences lacked the *–s* or *–ed* inflections. The rationale for including the responses for those sentences in scoring might be questioned at first glance. To justify inclusion of these in the data analysis, we should at first answer the question *why should they not be included?* It might be argued that they should be excluded because they might be misleading as that participants might have detected the errors in the presented stimuli but they repeated them as they were because they thought it was what they should do. Or, the stimuli were imitated verbatim in which case the response did not represent the participants’ linguistic competence. If this is why such responses should not be included in scoring, then there are more convincing factors that support their inclusion. First, participants were told to repeat sentences in correct English. Second, two of the five training trials were modeled by the researcher to show participants how their responses should have been when items were ungrammatical. Third, participants did spontaneously correct some of the sentences as Table 5.2 shows. This shows that the participants understood what they were required to do.

Table 5.2: Rate of correction in the ungrammatical sentences in the sentence elicited imitation task by Arab and Chinese participants (scores in parenthesis)

	Agreement	Past
Arab Low-level group	(15/41) 36.5%	(25/66) 37.8%
Arab Mid-level group	(25/61) 40.9%	(37/72) 51.3%
Arab High-level group	(36/45) 80%	(48/60) 80%
All Arab participants	(76/147) 51.7%	(110/198) 55.5%
Chinese Low-level group	(25/58) 43.1%	(31/85) 36.4%
Chinese Mid-level group	(27/59) 45.7%	(25/71) 35.2%
Chinese High-level group	(42/59) 71.1%	(50/73) 68.4%
All Chinese participants	(94/176) 53.4%	(106/229) 46.2%

However, two types of responses were excluded from the counting. The first type comprised verbs in contexts which were phonologically ambiguous (i.e. verbs followed by homophonic stops as in *travelled to*, or inter-dental fricatives as in *helped them*, or alveopalatal fricatives as in *wants some*). Commission errors, where a non-target-like inflection was used instead of the target inflection (i.e. *–ing* instead of *–s*, and *–ed*, or *–s* instead of *–ed*, or *–ed* instead of *–s*) constituted the second type of responses that were excluded.

5.4 Results

In the presentation of the results from the sentence elicited imitation task described in 5.3.4, three themes will be addressed. These are the accuracy, the developmental pattern and the L1 effects. First, accuracy refers to the correct production of past tense or verbal agreement morphology and this will be demonstrated in rates calculated from the number of correctly supplied morphological items out of the total number of morphological items in obligatory contexts. Second, the developmental patterns refer to the progress (or its lack) in the production of the same morphology that occurs between the same L1 learners at one L2 proficiency level and others at a higher level. These will be exhibited in trajectories showing the differences between Low, Mid and High proficiency groups. The third theme is L1 effects; this refers to the difference between the pair of L1 groups who are at the same proficiency level. This will be depicted in graphs showing the rates or scores of Chinese and Arabic proficiency groups.

All the data collected were coded into SPSS and analysed using relevant statistical analyses. The data were examined for non-normal distribution using Shapiro-Wilk test of normality. This revealed that non-normal distribution existed in the data of many groups (Arab and Chinese at different proficiency levels). Therefore, distributional assumptions of parametric tests were not met for many variables and consequently non-parametric analyses were used in line with standard procedures (Larson-Hall, 2010). Four types of nonparametric tests were used as follows: (1) Mann-Whitney Test: this is an independent samples test used to compare the means of two different groups. I used it here to compare, for example, Arab and Chinese participants at the same proficiency level. (2) Kruskal-Wallis Test: this is an independent samples test used to compare the means of three or more different groups. I used it here to compare, for example, the same-language groups at the three different proficiency levels to test for development with rising proficiency. When the result of Kruskal-Wallis was significant, Mann-Whitney was used as a post-hoc test (see (1) above). (3) Wilcoxon Signed Ranks Test: this is a paired-samples test used to compare two mean scores that come from the same group. I used it here to compare, for example, the results of the production of the inflection in cluster and non-cluster contexts by the same group of participants. (4) Friedman Test: this is a paired-samples test used to compare three or more mean scores that come from the same group. I used it here to compare, for example, the results of the production of the inflection in its four allomorphic variant contexts by the same group of participants. When the result of Friedman test was significant, Wilcoxon Signed Ranks Test was used as a post hoc test (see (3) above).

5.4.1 Past Tense Production: Arab vs. Chinese

As elaborated in section 5.3.4, the stimuli of sentence elicited imitation task contained 36 verbs in obligatory context for past tense. Table 5.3 reports the overall scores and percentages for all non-native groups.

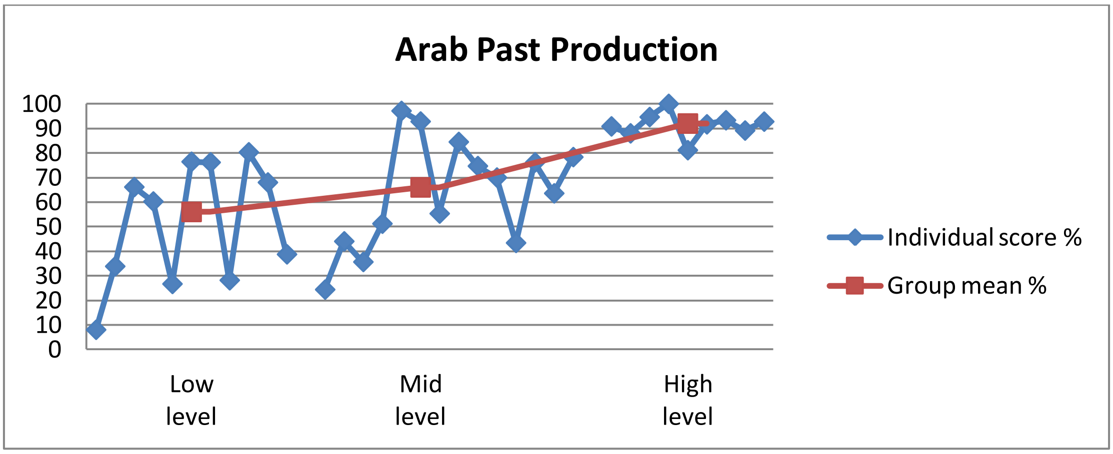
Table 5.3: Scores, percentages & standard deviations of past production by non-native groups (the number of participants in each group is in parentheses).

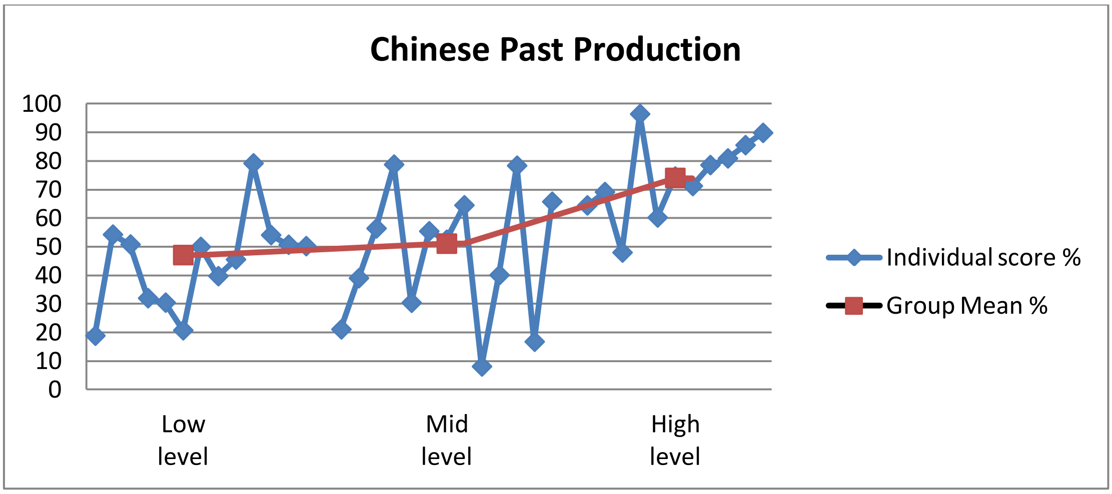
Arabic speakers				Chinese speakers			
	Score	%	SD		Score	%	SD
Low (n=11)	177/317	55.83	24	Low (n=13)	193/411	46.95	16
Mid (n=14)	276/416	66.34	21	Mid (n=13)	202/399	50.62	22
High (n=9)	275/297	92.59	5	High (n=11)	269/362	74.30	13

Table 5.3 shows that the production of past tense morphology was highly variable by Arab and Chinese participants at the Low and Mid proficiency levels. However, while the Arab High-level group produced the morphology accurately with little intra-group variation, its Chinese counterpart did so variably.

It is observed from Table 5.3 that the standard deviations (SD) of the results are large at Low and Mid levels of proficiency and get relatively smaller at the High level for both language groups. Larger SDs indicate that a wider distance between individual rates and the mean of the group results exists. Therefore further investigation of individual results is needed here in order to obtain a clearer picture of their performance. To illustrate the distance between individual rates and group means, Figure 5.1 visualises individual rates and group means. The same figure also shows trajectories of development with rising proficiency.

Figure 5.1: Past tense production by non-natives: Individual scores and group means





Large variation in individual results of participants at the same proficiency level can be observed for both Chinese and Arab learners alike in Figure 5.1. Such within group variation in the production of inflectional morphology is not unusual in SLA research. For example, Goad et al. (2003) reported that their Chinese participants who were at the same proficiency level (high intermediate/ low advanced proficiency range) in L2 English fell in two distinct groups as regards to their production of the verbal agreement morpheme. One group supplied the inflection on average 10% of the time and the other 49% of the time (Goad et al., 2003: 256). The researchers attributed this variation to L1-transferred prosodic constraints. They proposed that some learners realise that what their grammars in terms of prosody allow and what the L2 requires do not match and thus delete the inflection comprehensively. Others try to accommodate the L2 inflection according to what their grammars allow and the result is variability in production. The variation in Figure 5.1 here does not conform to this analysis because, as will be shown in section 5.4.5 below, the data shows no evidence for the production of the inflections being affected by prosodic constraints. The point here is that this variation is not unusual, but it nonetheless needs an interpretation. I will discuss this further in Chapter eight in the light of the results reported in this and the following two chapters.

It can also be seen in Figure 5.1 that there was an increase in the production of target-like past morphology with rising proficiency for both Arab and Chinese participants alike. To determine the significance of these developments across proficiency levels, Kruskal-Wallis tests were performed on Arab and Chinese learner results separately. Results from the Arab groups showed the following:

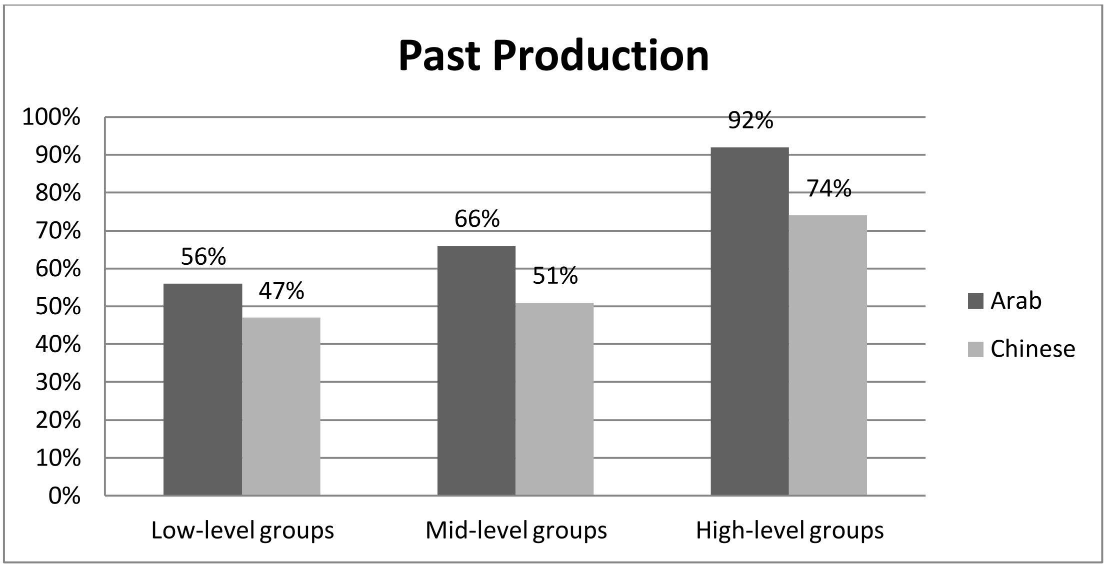
- There was a significant difference among Arab groups ($\chi^2(2)=15.825$, $p<.001$, $\eta^2=.47$). Post hoc testing (Mann-Whitney) indicated that the Low and Mid-level groups were not statistically different ($U=55$, $p=.228$) but that both performed significantly below the High-level group (Low vs. High: $U=00$, $p<.001$, $r=.84$; Mid vs. High: $U=15$, $p<.005$, $r=.63$)
-

Results from the Chinese groups showed the following:

- There was a significant difference among Chinese groups ($\chi^2(2)=13.174$, $p<.005$, $\eta^2=.36$). Post hoc testing (Mann-Whitney) indicated that the Low and Mid-level groups were not statistically different ($U=74$, $p=.590$) but that both performed significantly below the High-level group (Low vs. High: $U=13$, $p<.005$, $r=.69$; Mid vs. High: $U=22$, $p<.005$, $r=.58$)

To examine whether there was an L1 effect on the production of past tense morphology, a comparison between Arab and Chinese groups at the same proficiency levels was performed. Figure 5.2 illustrates the results presented in Table 5.3 above.

Figure 5.2: Past tense production: Arab and Chinese groups' comparison



It can be observed from Figure 5.2 that the Arab groups performed somewhat better than the Chinese groups at every proficiency level with the advantage increasing for the Arab groups at the highest level. To determine the significance of these differences between Arab and Chinese groups, Mann-Whitney tests comparing participants at the same proficiency level were performed. The results showed the following:

- The difference between Arabic and Chinese speakers in the production of past tense morphology was non-significant at the Low level (U=58, p=.434) and the Mid level (U=56, p=.094, r=.32), but the Arab participants performed significantly better than the Chinese ones at the High level (U=12, p<.005, r=.64).

To sum up the results on the production of past tense morphology, we see that while the Chinese participants produced morphology variably at the three levels of proficiency, the Arab participants did so only at the Low and Mid levels and supplied the morphology consistently (92%) at the High level. The developmental trajectories for Arab and Chinese groups were similar to a large extent as progress took place at every successive level of proficiency, but the development was statistically significant only at the High level for both groups. Moreover, the comparison between groups at the same proficiency levels demonstrated no statistical differences between Arab and Chinese participants at the Low and Mid levels but the Arab participants performed significantly better than the Chinese at the High level.

The results of the verbal agreement inflection production are presented next.

5.4.2 Verbal Agreement Production: Arab vs Chinese

As detailed in section 5.3.4, the stimuli of the sentence elicited imitation task contained 31 verbs in obligatory context for third-person *–s*. Table 5.4 reports the scores and percentages for all non-native groups.

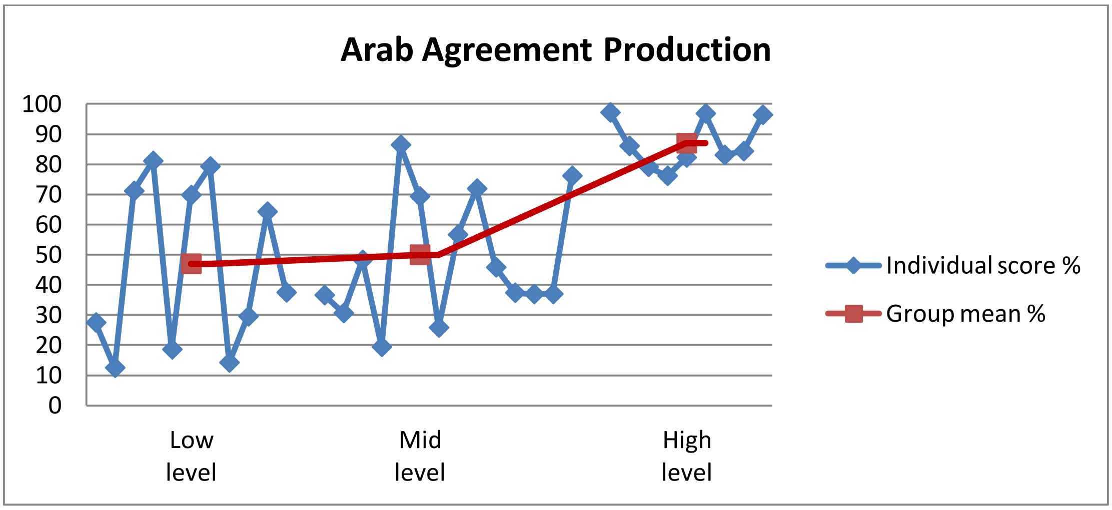
Table 5.4: Scores, percentages & standard deviations of verbal agreement production by non-native groups (the number of participants in each group is in parentheses)

Arabic Speakers				Chinese speakers			
	Score	%	SD		Score	%	SD
Low (n=11)	140/298	46.97	27	Low (n=13)	192/354	54.23	11
Mid (n=14)	170/342	49.70	20	Mid (n=13)	168/352	47.72	23
High (n=9)	226/260	86.92	8	High (n=11)	217/298	72.81	12

As can be seen here, the Arab and Chinese participants produced the *–s* inflection highly variably at the Low and Mid levels of proficiency and while the Arab High-level participants supplied the inflection consistently, their Chinese counterparts produced it variably.

Similar to past tense morphology production, the results of the verbal agreement production reveal large SDs from the mean value. To get a clearer picture of individual performance, Figure 5.2 depicts the individual results and their spread around group means. The same figure visualizes the developmental trajectories of Arab and Chinese groups based on the results presented in Table 5.4 in order to illustrate development in the production of the third-person *–s* inflection with rising proficiency.

Figure 5.3: Verbal agreement production by non-natives: Individual scores and group means.



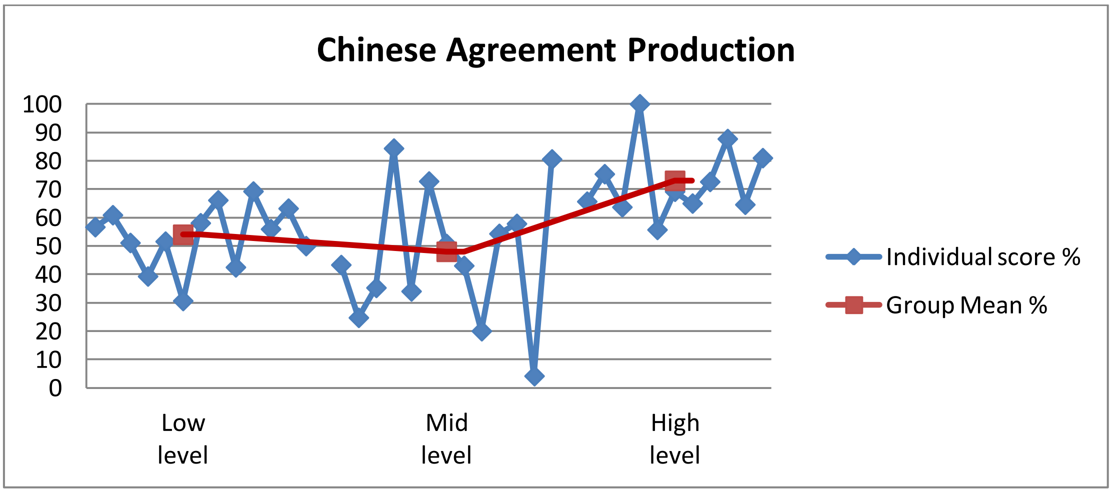


Figure 5.3 shows wide variations in scores of participants within the same proficiency group especially at Low and Mid levels. This variation, however, shrinks at the High level. This is consistent with the picture that arose from the past tense morphology production results. As noted above, this point will be discussed further in Chapter eight in the light of the results reported in this and the following two chapters.

Figure 5.3 also shows that no development seems to have occurred between the Low to Mid proficiency levels for either Arab or Chinese participants. Indeed there is an apparent dip of development at the Mid level for the Chinese group. However, a very marked progress appears to have taken place at the High proficiency level for both Arab and Chinese groups alike. To determine the significance of the development, Kruskal-

Wallis tests were performed on Arab and Chinese learner results separately. Results from the Arab groups showed the following:

- There was a significant difference among Arab groups ($\chi^2(2)=16.125$, $p<.001$, $\eta^2=.48$). Post hoc testing (Mann-Whitney) indicated that the Low and Mid-level groups were not statistically different ($U=69$, $p=.661$) but that both performed significantly below the High-level group (Low vs. High: $U=4$, $p<.005$, $r=.77$; Mid vs. High: $U=6$, $p<.001$, $r=.74$).

Results from the Chinese groups showed the following:

- There was a significant difference among Chinese groups ($\chi^2(2)=11.783$, $p<.005$, $\eta^2=.32$). Post hoc testing (Mann-Whitney) indicated that the Low and Mid-level groups were not statistically different ($U=65$, $p=.317$) but that both performed significantly below the High-level group (Low vs. High: $U=16$, $p<.005$, $r=.65$; Mid vs. High: $U=25$, $p<.01$, $r=.54$).

To examine whether there was an L1 effect on the production of third-person *–s*, a comparison between Arab and Chinese groups at the same proficiency level was performed. Figure 5.4 illustrates the results presented in Table 5.4 above.

Figure 5.4: Agreement production: Arab and Chinese groups' comparison

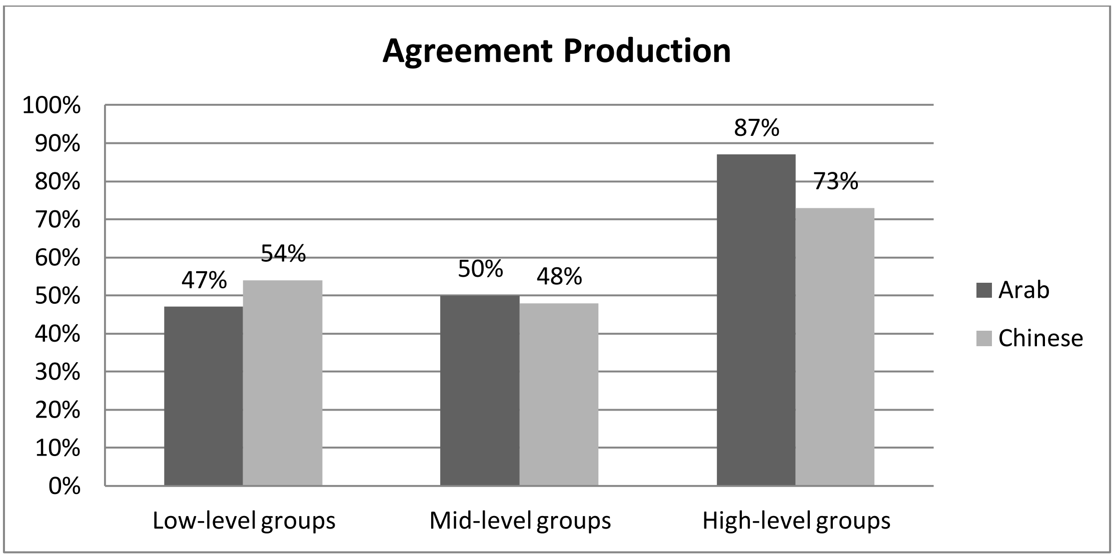


Figure 5.4 shows very small differences between the Arab and Chinese groups at the Low and Mid levels. At the High level, however, it demonstrates that the Arab group performed noticeably better than its Chinese counterpart as in the past tense results. To determine the significance of these differences, Mann-Whitney tests comparing Arab and Chinese groups at the three proficiency levels were performed. The results showed the following:

- The differences between the Arabic and Chinese speakers in the production of third-agreement *–s* were not significant at the Low level (U=64, p=.664) or the Mid level (U=88, p=.884), but the Arab participants performed significantly better than the Chinese ones at the High level (U=17, $p<.05$, r=.55).

In sum, the results of the production of the third-person *-s* showed that while the Chinese participants supplied the inflection variably at Low, Mid and High levels of proficiency, the Arab participants were variable only at the Low and Mid levels and produced the inflection more consistently at the High level. The developmental trajectories for Arab and Chinese groups were similar to a large extent as no progress took place between the Low level to Mid level but a remarkable progress occurred at the High level. Moreover, the comparison between the Arab and Chinese groups at the three proficiency levels demonstrated no statistical differences between any of the Low or Mid-level groups but the Arab participants at the High level were significantly better than their Chinese counterparts.

Overall, results of the production of past tense and verbal agreement morphology presented above showed similar symmetries in terms of variability, development and L1 effects. The variability appeared at the Low, Mid and High proficiency levels for Chinese participants and at the Low and Mid levels for Arab participants. The development in the production of morphology among participants at different proficiency levels was significant at only the High level for both Arab and Chinese participants alike. The L1 effect was statistically significant only at the High proficiency level in both types of morphology, giving an advantage of Arab participants over their Chinese counterparts.

I turn now to see if the variable use of the inflection and/or the observed differences between the Arab and Chinese participants can be associated with specific (L1-

transferred) phonological constraints. Therefore, the results presented above will be broken down by allomorph type, cluster type, prosodic structure, and (in case of past tense) the verb's (ir)regularity. I will start with allomorph and cluster types.

5.4.3 Past Production: Allomorph Type and Simple vs. Complex Codas

As shown before, the past tense inflection *–ed* has four allomorphs, which are (1) VV-d as in *played* /eid/, (2) CV-d as in *wanted* /ted/, (3) C-d as in *travelled* /ld/ and (4) C-t as in *walked* /kt/. The sentence elicited imitation task included equal number of verbs representing each of the four variants. For ease of reference I will separate the results by language group.

Table 5.5 presents the results of the production of the inflection in these four allomorphs by the Arab participants

Table 5.5: Production of past tense *–ed* by allomorph among Arab groups.

	CVV-d		CV-d		C-d		C-t	
Arabic	Score	%	Score	%	Score	%	Score	%
Low (n=11)	39/65	60	19/38	50	37/56	66.07	16/43	37.20
Mid (n=14)	69/91	75.82	38/57	66.66	53/78	67.94	23/52	44.23
High (n=9)	67/69	97.10	41/44	93.18	57/61	93.44	28/37	75.67

It can be observed here that variability extends to the production of the inflection in its four allomorphic variants especially in Low and Mid groups. Moreover, the production of the C-t allomorph is lower than that of other variants. To test for significance, Friedman tests were carried out on the rates for each proficiency group separately and the results showed the following:

- For the Arab Low-level group, there was a significant difference in the production of the past tense *–ed* across the four allomorphs ($\chi^2(3)$=10.469, p<.05, η^2=.22). To examine where the significant difference/s exactly occurred, post-hoc Wilcoxon signed rank tests were performed. These showed that the production of the C-t allomorph was significantly lower than that of the C-d allomorph (*Z*=-2.668, p<.01, r=.56) and there were no significant differences between the results of any of the other combinations.

- For the Arab Mid-level group, there was a significant difference in the production of the past tense *–ed* across the four allomorphs ($\chi^2(3)=8.951$, $p<.05$, $\eta^2=.16$). Post-hoc Wilcoxon signed rank tests showed that the production of the C-t allomorph was significantly lower than that of the CVV-d ($Z=-2.550$, $p<.05$, $r=.48$) and the C-d ($Z=-1.994$, $p<.05$, $r=.37$) and there were no significant differences between the results of any of the other combinations.

- For the Arab High-level group, there was no significant difference in the production of the past tense *–ed* across the four allomorphs ($\chi^2(3)=3.818$, $p=.282$).

To check if the phonological effects observed in the production of the inflection in its C-t allomorphic variant can be associated with a more general constraint on the production of consonant clusters, I combine the results of the VV-d and CV-d allomorphs and the C-d and C-t allomorphs together based on Table 5.5 and compare them with each other. Figure 5.5 gives the overall trend of these comparisons for the Arab groups.

Figure 5.5: The production of past tense *–ed* broken down by consonant cluster and non-cluster by the three Arab groups.

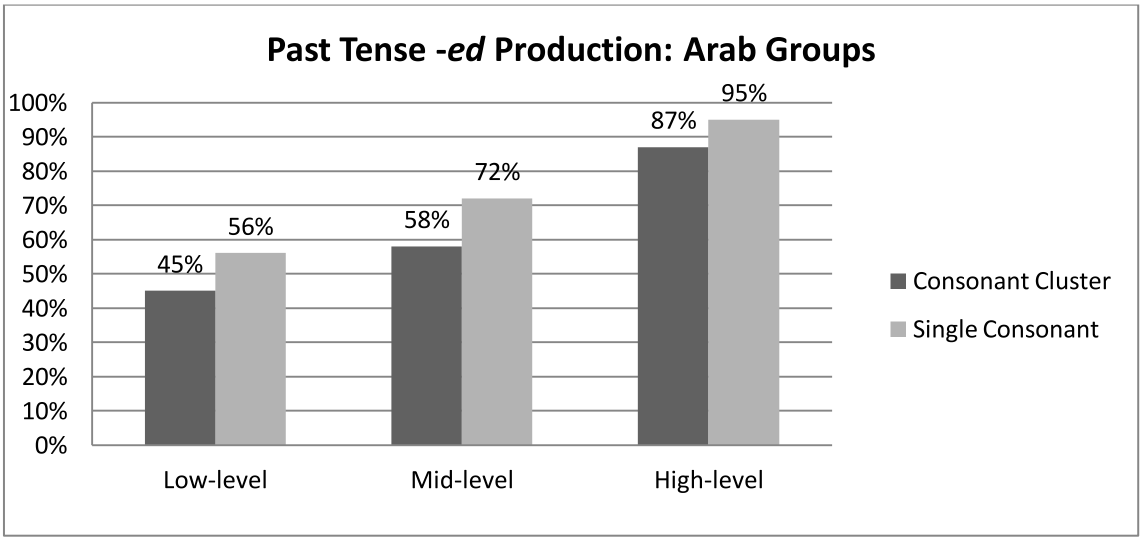


Figure 5.5 demonstrates that the production of the inflection is seemingly higher in verbs where it does not give rise to a consonant cluster especially in the performance of Mid- and High-level groups. To test for significance, Wilcoxon signed rank tests were performed on the results of each group separately. Results were as follows:

- The difference between the production of the inflection in cluster and non-cluster contexts was not significant in the performance of Arab participants at the Low level *(Z* =-.459, p=.646) and High level *(Z* =-1.612, p=.107), but at the Mid level, the production rates were significantly lower in verbs ending with consonant cluster than non-consonant cluster *(Z*=-2.197, p<.05, r=.41).

The Arab groups' results reported here demonstrated that there were some phonological effects on the production of the regular past tense morpheme. These effects appeared in the low rates of the production of the inflection in verbs ending with C-t structure by Low and Mid-level groups and the significantly higher rates of supplying the inflection in verbs not ending with consonant clusters compared to those ending with a consonant cluster in the performance of the Mid-level group. The High group's results, however, showed no phonological effects in production.

The Chinese participants' results of the production of the past tense *–ed* inflection across the four allomorphs are reported in Table 5.6.

Table 5.6: Production of past tense *–ed* by allomorph among Chinese groups.

	CVV-d		CV-d		C-d		C-t	
Chinese	Score	%	Score	%	Score	%	Score	%
Low (n=13)	39/82	47.56	22/56	39.28	30/69	43.47	27/68	39.70
Mid (n=13)	40/81	49.38	23/50	46	31/72	43.05	25/59	42.37
High (n=11)	57/83	68.67	35/45	77.77	47/67	70.14	38/53	71.69

Similar to the Arab group's results, the Chinese group's results presented in Table 5.6 show that variability extends to the production of the inflection in its four allomorphic variants. Friedman tests were carried out on the rates of production of the four allomorphs for each proficiency group separately and the results showed the following:

- There were no significant differences in the production of the past tense *–ed* across the four allomorphs in the performance of the Chinese participants at any level (Low: $\chi^2(3)$=.782, p=.854; Mid: $\chi^2(3)$=.838, p=.840; High: $\chi^2(3)$=.726, p=.867).

Similar to the analysis applied on the Arab groups' results concerning the comparison between the consonant cluster and non-cluster structures, I compare the same structures in the Chinese groups results through combining the results of C-d and C-t allomorphs on the one hand and those of the VV-d and CV-d allomorphs on the other based on Table 5.6. Figure 5.6 gives the overall results of these comparisons for the three Chinese different proficiency-level groups.

Figure 5.6: The production of past tense –ed broken down by consonant cluster and non-cluster by the three Chinese groups.

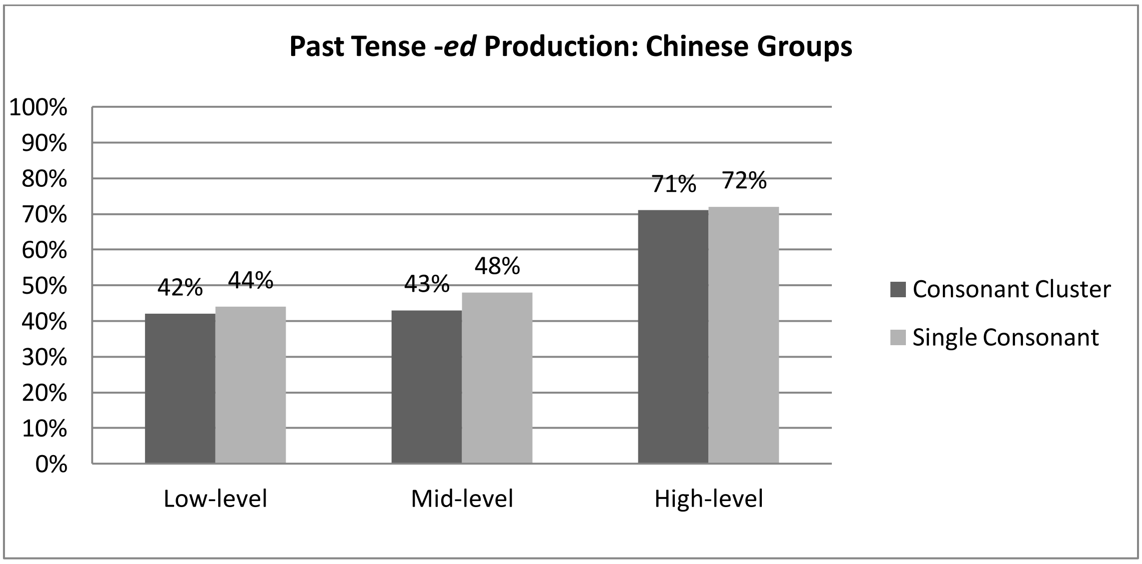


Figure 5.6 demonstrates that there are very few differences between the production of the inflection in consonant cluster and non-cluster contexts. To test for significance, Wilcoxon signed rank tests were performed. Results were as follows:

- There were no significant differences in the production of the *–ed* inflection between cluster and non-cluster contexts in the performance of Chinese participants at any level (Low: *Z*=-.175, p=.861; Mid: *Z*=-.863, p=.833; High: *Z*= -.051, p=.959).

The Chinese groups' results demonstrated no phonological effects on the production of the past tense *–ed* morpheme at any either Low, Mid or High level. The production rates of the inflection were similar in verbs of different word-final phonological structures.

Recall that my rationale for breaking down the results of the past *–ed* production by allomorph and coda types was to examine whether the variability at the Low and Mid levels by both Arab and Chinese participants and the Arab participants' advantage over their Chinese counterparts at the High level (described in 5.4.1 and 5.4.2) can be explained by specific L1 transferred phonological constraints. The assumption here is that if suppliance of the morpheme is affected by phonological transfer, we should see lower rates in the production of those allomorphic variants that involve consonant clusters word finally in the performance of the Chinese rather than Arab participants. However, although some phonological effects were detected in the performance of the Arab participants at the Low and Mid levels, no such effects were found in the Chinese participants' performance at any level. Therefore, based on these results, we can safely conclude that neither variability at Low and Mid proficiency levels nor differences between the Arab and Chinese High proficiency level groups can be explained by constraints on the shape of the coda in which the inflection occurs.

5.4.4 Agreement Production: Allomorph Type and Simple vs. Complex Codas

Now turning to the results on agreement, I test for possible phonological effects in line with procedures for past tense described above.

Similar to the past tense inflection -*ed*, the third-person *–s* creates four allomorphs, which are (1) VV-z as in *plays* /eiz/, (2) CV-z as in *loses* /zez/, (3) C-z as in *travels* /lz/ and (4) C-s as in talks /ks/, and the sentence elicited imitation task included equal number of verbs representing each of the four variants. Table 5.7 presents the results of the production of verbal agreement broken down by the four allomorphs for the Arab groups.

Table 5.7: Production of verbal agreement –s by allomorph among Arab groups.

	CVV-z		CV-z		C-z		C-s	
Arabic	Score	%	Score	%	Score	%	Score	%
Low (n=11)	31/72	43.05	33/58	56.89	38/74	51.35	38/94	40.42
Mid (n=14)	46/83	55.42	32/58	55.17	41/97	42.26	51/104	49.03
High (n=9)	58/62	93.54	44/53	83.01	57/67	85.07	67/78	85.89

It can be observed from Table 5.7 that variability was experienced by the Arab participants at the Low and Mid levels in the production of the agreement inflection in all of its four allomorphic variants. To test if any allomorph was easier or more difficult to produce by these participants, Friedman tests were carried out on the rates of production of the four allomorphs for each proficiency group separately and the results showed the following:

- There were no significant differences in the production of the third-person *–s* across the four allomorphs in the performance of Arab participants at any level (Low: $\chi^2(3)=4.806$, $p=.187$; Mid: $\chi^2(3)=2.530$, $p=.470$; High: $\chi^2(3)=1.338$, $p=.720$).

Having found no phonological effect on the type of allomorph leads us to check if there is a more general effect on the type of the cluster that morpheme occurs in. The C-z and C-s allomorphs are consonant clusters but the VV-z and CV-z allomorphs are not. So, based on Table 5.7, I combine the results of the allomorphs that create a consonant cluster and compare them with the results of allomorphs that do not create a consonant cluster word finally. Figure 5.7 demonstrates these comparisons.

Figure 5.7: The production of verbal agreement *–s* broken down by consonant cluster and non-cluster by the three Arab groups.

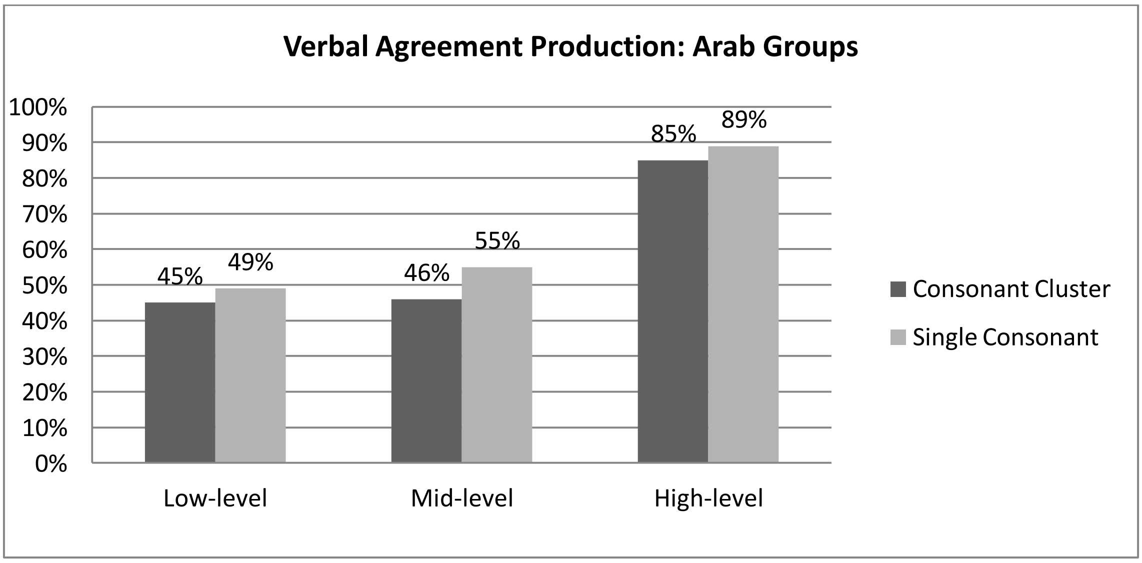


Figure 5.7 shows that the three Arab proficiency groups produced the inflection slightly higher in verbs ending with a no consonant cluster. To test for significance, Wilcoxon

signed rank tests were performed on the results of each proficiency group separately. The results were as follows:

- There were no significant differences between the production of the inflection in cluster and non-cluster forms in the performance of Arab participants at any level (Low: Z=-1.245, p=.213; Mid: Z=-1.350, p=.177; High: Z=-1.127, p=.260).

These results on allomorph and cluster types indicate that the production of the agreement morpheme by these Arab participants is not affected by phonological constraints on its shape. This leaves the reason for variability by Arab Low and Mid level participants unexplained.

I now turn to the Chinese participants' results of the production of the verbal agreement *–s* broken down by the four allomorphs to check for phonological effects. These are presented in Table 5.8.

Table 5.8: Production of verbal agreement *–s* by allomorph among Chinese groups.

	CVV-z		CV-z		C-z		C-s	
Chinese	Score	%	Score	%	Score	%	Score	%
Low (n=13)	40/79	50.63	36/68	52.94	46/92	50	70/115	60.86
Mid (n=13)	37/81	45.67	29/69	42.02	45/92	48.91	57/110	51.81
High (n=11)	44/66	66.66	52/62	83.87	55/83	66.26	66/87	75.86

Similar to the results of their Arab counterparts, the results of the Chinese participants in Table 5.8 shows that variability extends to the production of the inflection in its four allomorphic variants. To check for significance, Friedman tests were carried out on the rates of production of the four allomorphs for each proficiency group separately and the results showed the following:

- There were no significant differences in the production of the third-person *–s* across the four allomorphic variants in the performance of Chinese participants at any of the Low level ($\chi^2(3)$=1.320, p=.724) or Mid level ($\chi^2(3)$=4.887, p=.180). However, the results of the High level group showed a significant difference ($\chi^2(3)$=7.898, p<.05, η^2=.18). Post-hoc Wilcoxon signed rank tests showed that the production of the inflection in verbs ending with the CV-z structure was significantly higher than that of the VV-z structure (*Z*=-2.310, $p<.05$, r=.49) and there were no significant differences between any of the other combinations.

Further, to examine if there is a more general phonological constraint on the type of the cluster where the agreement inflection occurs, I compare the production rates of the inflection in verbs ending with a consonant cluster (C-z and C-s allomorphs) with these in verbs not ending with a consonant cluster (VV-z and CV-z allomorphs) based on the results presented in Table 5.8. Figure 5.8 shows these comparisons.

Figure 5.8: The production of verbal agreement –*s* broken down by consonant cluster and non-cluster by the three Chinese groups.

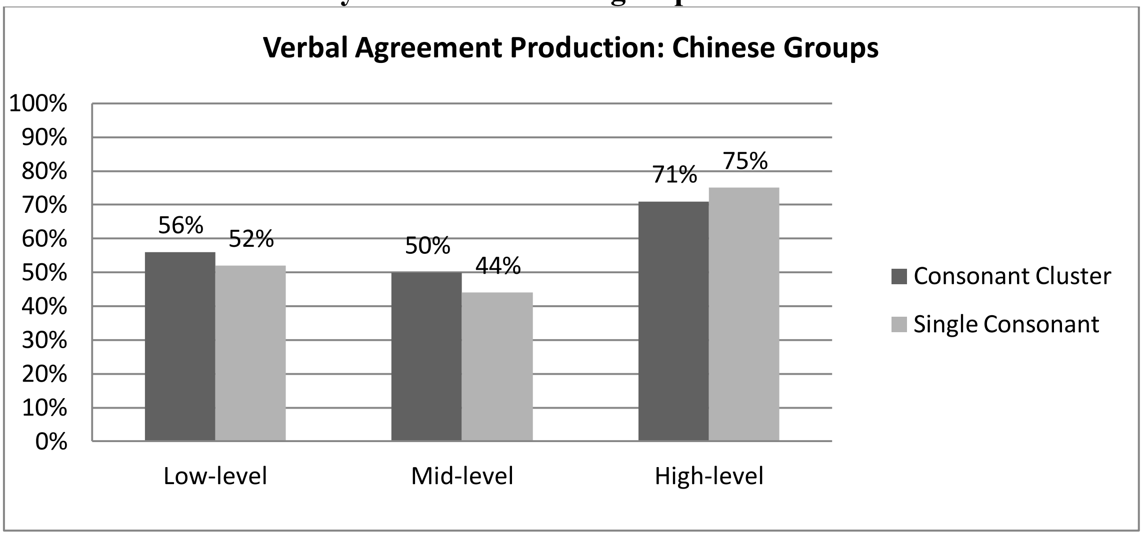


Figure 5.8 shows that there are very few differences between the production of the inflection in cluster and non-cluster contexts. To test for significance, Wilcoxon signed rank tests were performed. The results were as follows:

- There were no significant differences between the production of the inflection in cluster and non-cluster forms in the performance of Chinese participants at any level (Low: *Z*=-.454, p=-.650; Mid: *Z*=-1.726, p=.084; High: *Z*=-.765, p=.444).

The Chinese groups' results indicate that there are no phonological effects on the production of the verbal agreement –*s* morpheme at any of the Low, Mid or High levels. The production rates of the inflection were similar in verbs of different word-final phonological structures. One exception to this is the result showing that the production of the VV-z allomorph was significantly lower than that of the CV-z allomorph in the performance of the Chinese High level group. Yet, this is not the type of effect that would be expected if a constraint prohibiting consonant clusters was at play.

All in all, similar to the results of the past *–ed* production, the results of the verbal agreement *–s* broken down by allomorph and coda types indicate that neither variability at Low and Mid proficiency levels nor differences between the Arab and Chinese High proficiency level participants can be explained by constraints on the shape of the coda in which the inflection occurs.

I turn now to check if variability or L1 effects can be explained by another type of (L1-transferred) phonological constrains, namely prosodic constraints.

5.4.5 Past and Agreement Production: Prosodic Structures

The results of past and verbal agreement are further broken down by the type of the prosodic structure in which the inflection occurs. As elaborated in section 5.3.5, the prosodic structures I target are: (1) inflection as onset, (2) inflection as coda, (3) inflection as foot internal and (4) no option for the inflection within the PWd. In the first three types, the inflection is assumed to be easier to produce because it can be prosodified word-internally without violating specific L1 transferred constraints (Goad et al. 2003). I present the results of past tense *–ed* first and then move on to verbal agreement *–s*.

Tables 5.9 and 5.10 present the results of Arab and Chinese groups, respectively.

Table 5.9: Past production by prosodic structure by Arab participants

	Arab Low		Arab Mid		Arab High	
Past as onset	17/31	55%	23/46	50%	32/40	80%
Past as coda	19/38	50%	38/57	67%	41/44	93%
Past as foot Internal	74/124	60%	116/165	70%	117/124	94%
No option for past within PWd	1/7	14%	6/10	60%	3/3	100%

Table 5.10: Past production by prosodic structure by Chinese participants

	Chinese Low		Chinese Mid		Chinese High	
Past as onset	11/35	31%	16/39	41%	24/37	65%
Past as coda	22/56	39%	23/50	46%	35/45	78%
Past as foot Internal	78/168	46%	77/159	48%	114/153	74%
No option for past within PWd	7/16	44%	3/12	25%	4/13	31%

As seen in Tables 5.9 and 5.10, the number of contexts in which the inflection cannot be prosodified within the PWd is small, which does not enable us to perform a reliable comparison between the production of the inflection in such a context and other contexts where the inflection can be prosodified word internally. However, it is clear from both tables that past morphology was subject to variability in all contexts where it can be prosodified word internally. These structures are assumed to be easier than other contexts. However, this is not the attested pattern here. Both Arab and Chinese participants dropped the past *–ed* from such contexts.

I turn now to the verbal agreement production across the four prosodic structures. Table 5.11 presents the Arab groups' results.

Table 5.11: Agreement in production by prosodic structure by Arab participants

	Arab Low		Arab Mid		Arab High	
Agreement as onset	0/0	0%	0/0	0%	0/0	0%
Agreement as coda	33/58	57%	32/58	55%	44/53	83%
Agreement as foot Internal	71/162	44%	101/192	53%	124/140	89%
No option for agreement within PWd	36/78	46%	37/92	40%	58/67	87%

As this table shows, the Arab participants' data lacks any context where the inflection can be prosodified as onset. Friedman tests were carried out on the rates of production of the three remaining prosodic contexts for each proficiency group separately and the results revealed the following

- There were no significant differences in the rates of the verbal agreement inflection production in different prosodic structures for any of the Arab proficiency groups (Low: $\chi^2(2)= 2.390$, p=.303; Mid: $\chi^2(2)= .691$, p=.708; High: $\chi^2(2)= .467$, p=.792).

These results indicate that the verbal agreement inflection was produced variably regardless of the prosodic structure in which it occurred. Let us see if the same pattern occurs for Chinese participants.

Table 5.12 presents the Chinese participants results.

Table 5.12: Agreement production by prosodic structure by Chinese participants

	Chinese Low		Chinese Mid		Chinese High	
Agreement as onset	0/0	0%	0/1	0%	0/1	0%
Agreement as coda	36/68	53%	29/69	42%	52/62	84%
Agreement as foot-internal	113/194	58%	104/200	52%	108/153	70%
No option for agreement within PWd	43/92	47%	35/82	43%	57/82	69%

The agreement as onset context is also excluded here because only two cases were encountered in the whole dataset. Friedman tests were carried out on the rates of production of the three remaining prosodic contexts for each proficiency group separately and the results showed the following:

- There were no significant differences in the rates of the verbal agreement inflection production in different prosodic structures for any of the Chinese Low-level group ($\chi^2(2)=2.327$, $p=.312$) or Mid-level group ($\chi^2(2)=4.154$, $p=.125$). A significant difference however was found in the results of the Chinese High-level group ($\chi^2(2)=6.000$, $p=.050$, $\eta^2=.17$). Post hoc testing (Wilcoxon) showed that the production of agreement as coda was significantly higher than that of the agreement as foot-internal ($Z=-2.090$, $p<.05$, $r=.44$) but no significant difference was detected between the no option for agreement within PWd context and any of the other two contexts.

These results also show that the Chinese participants produced the inflection variably regardless of the prosodic structure in which it occurred. Although the production was higher where the inflection can be prosodified as coda, this does not seem to be due to the fact that it can be prosodified word-internally. This is because its production was higher than the production of the inflection in a word-internal context (agreement as foot-internal) but not word-external context (no option for agreement within PWd).

Therefore, these results clearly indicate that neither the variability in the production of past and agreement morphology by Arab and Chinese participants nor the difference between these L1 groups can be associated with L1-transferred prosodic constraints.

So far, no influence of phonological constraints has been detected in the data. However, such results do not exclude the possibility of a general phonological effect on word-final consonants. Therefore, the comparison between the production of regular and irregular past tense morphology provides a window to further investigate any effect for phonological constraints on the production of past tense morphology. These comparisons are presented hereafter.

5.4.6 Past Production: Regular vs. Irregular

The sentences of the elicited imitation task contained thirty-six obligatory contexts for past tense out of which 10 were irregular verbs and the rest were regular ones. Table 5.13 reports the results of the Arab groups

Table 5.13: Regular vs irregular past tense verb morphology production by Arab groups

	Regular			**Irregular**		
Arabic	Score	%	SD	Score	%	SD
Low (n=11)	111/202	54.95	27	66/115	57.39	23
Mid (n=14)	183/278	65.82	24	93/138	67.39	17
High (n=9)	193/211	91.46	7	82/86	95.34	6

Table 5.13 shows that variability extends to both types of verbs. There are very few differences in their production by the three Arab proficiency groups. To test for significance, Wilcoxon signed rank tests were performed. Results were as follows:

- There were no significant differences between the production of the past tense morphology in regular and irregular verbs in the performance of Arab participants at any level (Low: Z=-1.156, p=.248; Mid: Z=-.910, p=.363; High: Z=-1.400, p=.161).

The Chinese participants' results are presented in Table 5.14.

Table 5.14: Regular vs irregular past tense verb morphology production by Chinese groups

	Regular			Irregular		
Chinese	Score	%	SD	Score	%	SD
Low (n=13)	118/275	42.90	19	75/136	55.14	15
Mid (n=13)	119/262	45.41	26	83/137	60.58	16
High (n=11)	177/248	71.37	16	92/114	80.70	9

Similarly, Table 5.14 shows that both types of verbs were subject to variability in the performance of the Chinese participants at the three proficiency levels. Moreover, a clear pattern in the results of the three proficiency groups is that the suppliance of past tense was higher on irregular verbs than regular verbs. To determine the significance of the differences between these results, Wilcoxon signed rank tests were performed. Results were as follows:

- The production of past tense morphology in irregular verbs was significantly higher than in regular verbs in the performance of Chinese participants at the Low level (Z=-2.062, p<.05, r=.40) and Mid level (Z=-2.271, p<.05, r=.44), but at the High level, the difference was not statistically significant (Z=-1.511, p=.131).

To sum up, the results in this section show that the production of past tense in regular and irregular verbs was the same for Arab participants at all proficiency levels but it was lower in the former than the latter in the performance of Chinese participants at the Low and Mid levels.

5.5 Summary and Hypothesis Evaluation

In this section, the results will be summarised first. Then, the experimental hypotheses will be evaluated against the findings.

The present experiment has examined the phenomenon of variability, developmental patterns and L1 effects in the production of past tense and verbal agreement morphology. It was found that morphological variability appeared at Low, Mid and High proficiency levels in the performance of Chinese participants, whereas Arab participants experienced this phenomenon only at Low and Mid levels, overcoming it at the High level. The developmental patterns were the same across Arab and Chinese groups as no development was observed between the Low to Mid levels and a significant improvement took place at the High level. The L1 effect emerged only at the High level of proficiency with the Arab participants outperforming their Chinese counterparts. Pursuing the source of both variability and the difference between Arab and Chinese High level groups, the results were broken down by (1) allomorphic variant (VV-d/z, CV-d/z, C-t/s and C-d/z), (2) cluster shape (consonant cluster and non-cluster), (3) prosodic structure (PWd internal and PWd external analyses of the inflection) and (4) verb type (regular and irregular). The results of these analyses showed that

(1) Allomorphic variants: the production of all allomorphs of past and agreement inflections was variable by all groups except Arab High level participants. Disparity in their rates was found in three instances, which are a) Arab Low level participants produced C-t allomorph lower than C-d allomorph, b) Arab Mid level participants produced C-t allomorph lower that C-d and CV-d allomorphs, and c) Chinese High level participants produced the CV-z allomorph higher than the VV-z allomorph

(2) Cluster shape: The production of past and agreement inflections in cluster and non-cluster contexts was the same by all groups except the Arab Mid level group, who produced the past tense inflection higher in single consonant cluster codas than consonant cluster ones.

(3) Prosodic structure: all prosodic structures were found variable in the production of both past tense and verbal agreement. Comparison of cases where the inflection can and cannot be prosidified word internally failed for past tense because the number of contexts in the latter was very low. This comparison for verbal agreement showed no difference in the production of the inflection.

(4) Verb type: production of past was similarly variable in regular and irregular verbs by all groups except the Arab High level group. Differences in production rates appeared in two occasions where the Chinese participants at Low and Mid levels expressed past higher in irregular than regular verbs

The aim of the present experiment was to answer the main research question "*what is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?*" Six hypotheses were formulated and thus tested in the experiment. For convenience, these hypotheses will be repeated below. Each will be followed by conclusions.

> H1: *The production of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, and both language groups will improve similarly with rising proficiency.*

This hypothesis (based on OG) is partially confirmed. We have seen that Arab and Chinese participants experienced variability at lower proficiency levels and significant improvement in the production of morphology under scrutiny with rising proficiency was observed in the performance of both language groups. However, at the High proficiency level, while the Arab participants overcame variability, their Chinese counterparts did not. Therefore, it can be concluded here that this hypothesis is widely supported by the overall results, but it is challenged by the Arab participants' advantage over their Chinese counterparts at the High level.

> H2: *The production of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, but Arab participants will improve faster than their Chinese counterparts with rising proficiency.*

This Hypothesis (based on the MSBH) is confirmed. The attested variability at lower proficiency levels and the advantage that the Arab participants had over their Chinese counterparts at the High level are fully compatible with this hypothesis.

> H3: *Arab participants will have an L1-based advantage over their Chinese counterparts in the production of morphology from lower levels onwards.*

This hypothesis (based on the RDH) is partially supported. Arab participants had a significant advantage over the Chinese participants at the High proficiency level and this advantage was not due to (L1-transferred) phonological constraints affecting the performance of Chinese participants, which is compatible with this hypothesis.

However, the evidence that this advantage did not occur early on seems at odds with this hypothesis, posing a challenge for it.

> H4: *The production of morphology by Arab and Chinese participants at the same L2 proficiency level will be similar.*

This hypothesis (based on MSIH) is partially supported. The Low and Mid proficiency level participants from the two L1 backgrounds performed similarly showing similar rates of the production of past and agreement morphology. This is consistent with this hypothesis. However, both language High level groups showed significant differences in their production rates and these were not found to be due to phonological effects, an issue that poses a challenge for this hypothesis.

> H5: *The production of morphology by Arab and Chinese participants at the same L2 proficiency level will be similar and the production of morphology will be higher on verbs where the inflection can be prosodified word-internally than on verbs where it cannot.*

This hypothesis (based on the PTH) was not supported. The prosodic structure analysis did not show the expected patterns. For past tense *–ed*, the number of contexts where the inflection cannot be prosodified PWd internally was very small. Although this prevented the comparison with other contexts, it was clear from the results that the production of the inflection where it can be prosodified PWd internally was highly variable. For the verbal agreement *–s*, enough contexts for the two prosodification fashions existed. Comparisons here did not show any differences. Thus, the PTH was not supported by these results.

> H6. *Arab participants will have an L1-based advantage over their Chinese counterparts, who will produce the morphology higher on non-cluster than consonant cluster verbs and irregular than regular verbs.*

This hypothesis (based on PhRH) is not supported by the results reported in this chapter. First, no L1 effect was observed in the performance of Arab and Chinese participants at Low and Mid levels as both performed similarly variably in the production of past and agreement morphology. Second, at the High level, the Arab participants did have a significant advantage over their Chinese counterparts, but the suppliance was higher neither in no cluster contexts nor on irregular verbs.

One result that seems consistent with hypotheses five and six above is that the production of past was higher on irregular than regular verbs by Chinese Low and Mid level groups. Nevertheless, this performance does not seem to be due to phonological constraints on the production of the regular inflection and, moreover, even if it really is due to such constraints, this cannot account for variability or the advantage of Arab participants over their Chinese counterparts. This is due to the following reasons. First, even though the production of irregular past is significantly higher than the regular inflection by Chinese Low and Mid level participants, the production rates of past on irregular verbs are 55% for the Low-level group and 61% for the Mid-level group and such rates do still show robust variability. If variability is due to phonological constraints on the production of the regular morpheme, higher rates in the production of irregular morphology should have been attested. Second, the advantage of Arab participants over their Chinese counterparts appears only at the High level of proficiency. However, at this proficiency level, the Chinese participants' production of regular and irregular morphology is not statistically different. Therefore, the higher rates of past tense morphology production by Arab participants at the High level cannot be explained by the Chinese participants' performance being affected by phonological constraints.

In Chapter eight, the results from the present chapter will be discussed further in the light of the literature reviewed in Chapters two and three.

In the following chapter, we will see how these participants perceived the same morphological items. This will allow for a better understanding of the performance of these participants.

Chapter 6. The Perception of Past Tense and Verbal Agreement: The Study Design and Results

6.1 Introduction

This chapter reports on a cross-sectional study conducted on the perception of past tense and verbal agreement morphology by Arabic or Chinese adult L2ers of English. It was evident from Chapter five that the Arab and Chinese research participants experienced severe variability in the production of the morphology under investigation at Low and Mid proficiency levels. At the High proficiency level, however, while Arab participants were found consistent in their production, their Chinese counterparts were variable. The present chapter examines whether the observed morphological variability in the production of these participants extends to their perception. The aim is again attempting to locate the source of morphological variability in SLA.

As noted in Chapter three (section 3.2.2), I use the word perception in this research to refer to its two aspects. These are (1) phonological decoding and (2) sound interpretation. Chapter three (section 3.2.2) reviewed studies on the perception of past tense and verbal agreement, which showed that L2ers had perceptual difficulties. It was not clear in those studies whether poor phonological decoding abilities or a problem in the interpretation of the inflection was the source of the difficulty (Johnson and Newport, 1989; McDonald, 2000; Solt et al., 2004; McDonald 2006; Chen et al., 2007; McDonald and Roussel, 2010). The assumption was that a problem in the interpretation, but not in the phonological decoding, would more likely be caused by non-target syntactic representations. In this thesis, an experiment testing the perception and processing of past tense and verbal agreement morphology was designed. The perception part is dealt with in this chapter and the processing part is left to the next chapter (Chapter seven). A computerised-picture choice task was used for this objective, and the test stimuli, as will be shown in detail below, were controlled in a way as to allow examining whether a perceptual difficulty is caused by an inflection interpretation or phonological perception problems.

This chapter is organised as follows. Section 6.2 presents the research hypotheses. The methodology is outlined in section 6.3. The results are presented in 6.4 followed by the conclusion in which the research hypotheses are evaluated against the reported results in 6.5.

6.2 Research Hypotheses

To answer the main research question *What is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?,* five experimental hypotheses on the perception of these morphological items have been formulated. The first four hypotheses are parallel to the first four hypotheses presented in Chapter five (section 5.2) and are based on the assumption that perception mirrors production as the same syntactic representations are used in both modalities. Recall that the first four hypotheses in Chapter five are based on Organic Grammar (OG), the Modulated Structure Building Hypothesis (MSBH), the Representational Deficit Hypothesis (RDH) and the Missing Surface Inflection Hypothesis (MSIH), respectively. The fifth experimental hypothesis here is set up based on the findings of previous research reviewed in Chapter three (section 3.2.2), which showed that difficulties in perception might be caused by difficulties in the phonological decoding of inflections in complex codas.

First, OG (see Chapter two, section 2.3.3.1) views morphological variability as a phenomenon occurring because of a temporary absence of the syntactic representations, which are built incrementally during linguistic development based on the interaction between the input and fully available UG. L2ers' L1 is assumed to play no role in the building up of functional categories including those corresponding to past tense and verbal agreement functional morphology. Based on this proposal, Hypothesis one in Chapter five (section 5.2) was formulated and tested in the production experiment in the same chapter. The hypothesis was supported by the variability in the performance of Arab and Chinese Low and Mid level participants and the developmental patterns across Low, Mid and High level participants from both language backgrounds. However, the advantage of Arab participants at the High level was a challenge for this hypothesis. Now, testing this hypothesis in perception, it is assumed that perception mirrors production because both use the same syntactic representations. Therefore, the variability observed in production and the developmental patterns attested across different language and proficiency participants should be reflected in perception. This leads to the formulation of Hypothesis one based on OG.

> *H1. The perception of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, and both language groups will improve in accuracy similarly with rising proficiency.*

Secondly, similar to OG, the MSBH (see Chapter two, section 2.3.3.2) maintains that it is a temporary lack of syntactic representations that causes morphological variability. Different from OG, however, it proposes that the building up of L2 functional categories is affected by the presence or absence of these representations in the L1. Their presence will precipitate their acquisition. Based on this as well as on the similarities and difference between the language backgrounds of the participants, Hypothesis two in Chapter five (section 5.2) was formulated and tested in the production experiment in the same chapter. The hypothesis was widely supported in the variability in Low and Mid levels, the developmental patterns of the two language groups across three levels of proficiency and the advantage of Arab participants over their Chinese counterparts at the High level. As for perception, similar performance is expected if this hypothesis is on the right track. Hence, hypothesis two is formulated as follows.

> *H2. The perception of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, but Arab participants will improve faster than their Chinese counterparts with rising proficiency.*

Thirdly, the RDH, which proposes that uninterpretable features cannot be acquired in adult SLA if they are absent from previous linguistic experience(s), was the motivation for Hypothesis three in Chapter five (section 5.2). It predicts Arab participants to have advantage over their Chinese counterparts from early on in the production of past and agreement morphology because both features exist in Arabic but are absent from Chinese. Indeed, we observed an advantage for Arabs at the High level. No difference, however, was detected at Low and Mid levels, which was a challenge for this hypothesis. Now, for perception a similar hypothesis is formulated based on the RDH as follows.

H3. Arab participants will have an L1-based advantage over their Chinese counterparts in the perception of morphology from lower levels onwards.

Fourthly, experimental hypothesis four in Chapter five (section 5.2) was set up based o the MSIH (see Chapter two, section 2.3.3.4), which maintains that the syntactic structure is fully available in SLA from early on and the observed variability is caused by a processing failure during production. This hypothesis predicted that Arab and Chinese participants would perform similarly at all levels of proficiency unless phonological effects generated differences. This also was partially supported in the production data. Both Arab and Chinese participants at the Low and Mid levels performed similarly as expected. However, at the high level, Arabs did better than the Chinese and phonology was found not to be involved, which was a challenge for this hypothesis. Now, for perception, under the MSIH, no perceptual problems are expected to arise. This is for two reasons. First, syntactic structures relevant for past tense and verbal agreement morphology are assumed to be present in L2ers' underlying grammars from early on. Secondly, the results of the production of morphology (as reported in Chapter five show that the participants in this research have knowledge of the surface forms of these properties. Therefore, if morphological variability is caused by a processing failure due to communicative pressure during production, the perception should not be affected at all. This leads to the formulation of hypothesis four below.

H4. The perception of morphology will be accurate from early levels onwards by both Arab and Chinese groups alike.

Finally, the last experimental hypothesis that will be tested in this chapter pertains to the phonological decoding of the inflection. Lardiere (1998a: 21) maintains that the occurrence of the English inflection in consonant cluster codas does not only pose a difficulty in its production but also possibly in its perception, especially to L2ers whose L1s disallow such clusters word finally (e.g., Chinese). A study reviewed in Chapter three (section 3.2.2) by Solt et al. (2004) revealed that the perception of regular past tense inflection was difficult for their participants particularly in word-final consonant cluster contexts. The participants of Solt et al.'s study were native speakers of a variety of L1s including Mandarin Chinese, French, Arabic and Spanish and thus this did not inform whether L1 transfer was involved. Therefore, I take this further in the present experiment to examine if the perception of the inflection is affected by phonological constraints on its shape in consonant cluster codas and, moreover, whether this is due to an L1 transfer in this domain. Investigating L1 transfer here is possible as the participants are native speakers of two languages which contrast in allowing or disallowing consonant clusters in codas (Arabic and Chinese, respectively). This leads to the formulation of hypothesis five as follows:

H5. The perception of the inflection by the Chinese, but not the Arab, participants will be lower in cluster than non-cluster codas (in past tense and agreement contexts) and regular than irregular verbs (in past tense contexts).

6.3 Method

6.3.1 Design

This experiment tests the perception of English past tense and verbal agreement morphology by Arabic and Chinese speakers at three proficiency levels of L2 English- Low, Mid and High. For this purpose, a computerised picture choice task was used. In this task, participants were presented with sets of three pictures accompanied by aurally presented sentences and they were asked to choose the picture they considered best depicted by the oral stimuli. The choice of a picture indicated whether participants perceived the inflection or not. The oral stimuli included verbs which created various phonological structures word finally when the inflection was added to them.

6.3.2 Participants

The participants of this study are the same non-native participants who took part in the production study in addition to the group of 10 native speakers of British English who served as controls. All participants' background information and the non-native proficiency details are presented in Chapter four, sections 4.3.1 and 4.4, respectively. The reader is referred to those sections for more information.

6.3.3 The Computerised Picture Choice Task: Theoretical Assumptions

The present experiment uses a computerised picture choice task to test the perception of past tense and verbal agreement morphology in SLA of English. It is clear from the research reviewed in Chapter three (section 3.2.2) that no study in the field of SLA has used the picture-choice method to investigate the acquisition of English past and agreement morphology. Hence, prior to presenting the task used in this study, I briefly introduce the method and the theoretical assumption underlying its use.

In a picture choice task, participants are presented with a visual or auditory sentence and an array of two or more pictures from which they are asked to pick the one that best matches the sentence. Three types of data can be obtained from such a task. These are (1) the behavioral response (i.e., choosing a picture), (2) reaction time and (3) eye movement. This chapter deals with the first type and the other two types are dealt with in the next chapter (Chapter seven)

The assumption underlying the use of this method is that it taps linguistic competence through comprehension. This procedure of sentence-picture matching is considered a comprehension measure, which is "used to establish whether learners are able to process specific linguistic features in the input" (Ellis and Barkhuizen, 2005: 16). It is held that these non-verbal responses reflect participants' comprehension or non-comprehension of the verbal stimuli presented to them. For example, Kempe and MacWhinney (1998) used a picture-choice task to tap L2ers' comprehension of case marking morphology in Russian and German. The participants were administered an auditory sentence and two pictures and asked to choose the picture that matched the sentence. While choice of the target picture (the correct choice) was analysed as being an indication of the participants' comprehension of the agent or object information inherent in inflectional morphology, choice of the wrong picture was interpreted as a demonstration of their non-comprehension or unawareness of the same linguistic information.

Similar to Kempe and MacWhinney (1998), I assume in the task of the present experiment that choosing the correct picture indicates successful comprehension and choosing the wrong pictures indicates lack of comprehension. However, there is one factor unique to the present investigation that needs careful consideration. This is the nature of the English past tense and verbal agreement inflections. As elaborated in

Chapter three (section 3.3.1), adding the *–ed* and *–s* inflections to verbs potentially makes four different phonological structures word finally, two of which create a cluster of consonants (e.g, asked /kt/ learned /nd/ & asks /ks/, learns /nz/). As shown in Chapter three (section 3.2.2), the inflection in such contexts, as compared to other contexts where the inflection does not create a cluster of consonants, causes a perceptual challenge for L2ers of English (Solt et al., 2004). Bearing this in mind, we can assume that choosing the wrong picture in a picture choice task testing past tense and verbal agreement morphology does not necessarily indicate a lack of comprehension because this might rather be due to a difficulty in the phonological decoding of the inflection in consonant cluster contexts. The design of the task of the current study (as will be shown below) and the interpretation of its results take this issue into consideration.

The following section presents the picture-choice task used in this study.

6.3.4 The Computerised Picture Choice Task: Materials

The present experiment used a computerised picture choice task to test the perception of past tense and verbal agreement morphology. This task consisted of picture-sentence trials administered to participants on a computer screen. In each trial, participants viewed three pictures and listened to an auditory stimulus, and were asked to choose the picture described by what they had heard. The procedure of administering three pictures, rather than two as Kempe and MacWhinney (1998) did, was to minimise the possibility of participants getting correct answers by chance. The participants' behavioral responses, reaction times and the eye movements were recorded.

The number of picture-sentence trials in the task was 88 in total. These were divided into 27 experimental, 39 distractor, 12 experimental-like filler and 10 training trials. We introduce these items here as pairs of trials because each pair shared the same pictures but with two different auditory stimuli, which resulted in 44 pairs divided as follows: (a) Fifteen pairs tested past tense regular and irregular morphology, (b) twelve pairs tested the third-person verbal agreement *–s,* (c) twelve were experimental-like pairs of fillers and (d) five were training pairs. The pairs in (a) and (b) are called experimental and each of these contained one experimental trial and one distractor trial. Each trial involved three pictures and an auditory stimulus. The three presented pictures were a target picture (correct choice), a competitor picture (possible but wrong choice) and a foil picture (unrelated picture). More information on each type of trials is provided below. I start with the past tense trials.

Past Tense Trials

The task included 15 pairs of trials especially designed to test the perception of past tense morphology. Designing such trials needed special care because tense cannot be easily depicted visually. To test the perception of past tense morphology in the present picture-choice task, the following two criteria needed to be satisfied:

1- The trial must contain two pictures depicting the same action/state and clearly showing that they occurred at two different times.
2- The trial must make it clear that the first action/state happened/finished at a time before now (past) and the other is happening at the moment of speaking (present).

In the picture-choice task used in this study, the first criterion was satisfied as follows.

1- Pictures were attached to clocks showing time (example below).

Each of the three pictures was attached to a clock showing the time of the action. Clocks reading 8:00, 9:00 and 10:00 o'clock were attached to the first, second and third pictures respectively. The action/state depicted in the 10:00 o'clock picture was the same as that of either the 8:00 or 9:00 o'clock pictures, satisfying the first criterion of two pictures depicting the same action/state at two different times. The second criterion was satisfied as follows

2- The context of time was introduced in the distractor trials preceding the trials testing past tense morphology.

All the distractor trials preceding past tense trials started with an aural stimulus saying "it's ten o'clock right now. Listen and choose" followed by the distractor sentence saying something similar to "X was doing something two hours ago" or "it happened one hour ago." In this way, participants' answers to such distractor trials could show whether or not they were aware of the time of the action and the idea of past vs. present.

Setting the context for past in this way, the experimental trials administered an aural sentence containing a verb marked for past tense and in those stimuli, past tense was expressed solely by the morphology marking the verb. The following is an example of one pair of trials:

(6.1)

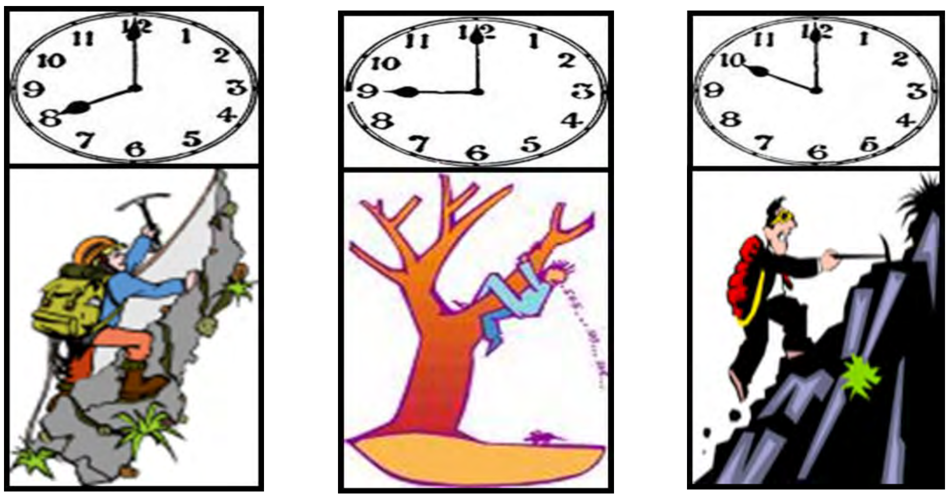
a: Distractor Trial

It's ten o'clock right now. Listen and choose:

"He was climbing something one hour ago"

b: Experimental Trial

"He climbed up the mountain"



As can be seen here, the distractor trial (6.1a) introduces the context of time. In response to this trial, choosing the 9 o'clock picture would indicate that the participant had understood the tense context. In the experimental trial (6.1b), the same pictures are presented but with an experimental sentence "he climbed up the mountain". I assumed here that successful perception of the past tense inflection on the verb would trigger participants to choose the 8 o'clock picture. Adversely, not perceiving the inflection would make participants confused between the 10 and 8 o'clock pictures, which would show up in the participants' behavioural response, reaction time and/or eye movements.

.

Past tense aural stimulus design: The experimental verbal stimuli testing past tense (15 sentences) ranged between four and eight words in length (mean=6.1). Three of these sentences contained irregular past tense verbs and the other 12 contained regular past tense verbs representing the four phonological variants that arise in a word final position when the *–ed* inflection is added (VV-d, CV-d, C-t and C-d) at a rate of three verbs from each form. Table 6.1 presents the verbs used in these sentences (for task sentences, see appendix F).

Table 6.1: All verbs used in the stimuli testing past tense morphology in the computerised picture-choice task

VV-d	CV-d	C-t	C-d	Irregular
Played	Painted	Walked	Climbed	Went
Stayed	Waited	Typed	Cleaned	Broke
Cried	Lifted	Watched	Phoned	Kept

Verbal Agreement Trials

The task also included 12 pairs of experimental trials especially designed to test the perception of third-person verbal agreement *–s*. In a present tense context, the form of the English verb (bare or inflected with *–s*) determines the number of the preceding noun (plural or singular). Using aural subject-drop sentences, experimental trials examined whether participants could recover the subject number depending solely on the number agreement marking on the verb. In the visual stimuli presented in these trials, two pictures (the target and the competitor) depicted a similar action performed by different number of subjects. The target picture depicted an action being performed by one character and the competitor picture depicted the same action but being performed by more than one character. A foil picture was also added. The following is an example of one pair of trials.

(6.2)

a: Experimental Trial

"Every party [….] dances happily"

b: Distractor Trial

"They are jumping all together"

Example 6.2 shows that the same pictures were introduced twice but with two different verbal stimuli. Two of the pictures depict the action of dancing. The difference is in the number of characters in each. In the experimental verbal stimuli "every party [….] dances happily", the subject of the sentence is not provided. I assumed that successful perception of the *–s* inflection on the verb 'dances' would trigger participants to choose the target picture (the single character dancing). Failure in perception, on the other hand, is assumed to lead participants to choose the competitor picture (two people dancing) or make them confused between the target and competitor pictures, which would also show up in the participants' behavioural response, reaction time and/or eye movements.

Verbal agreement aural stimulus design: The verbal stimuli in the 12 verbal agreement trials ranged between four and seven words in length (mean=5.8). The subject of the sentence was masked by silence. These stimuli contained inflected verbs representing the four phonological variants that arise in a word final position when the *–s* inflection is added (VV-z, CV-z, C-s and C-z) at a rate of 3 verbs from each form. Table (6.2) presents the verbs used in these sentences (for task sentences, see appendix F).

Table 6.2: All verbs used in the stimuli testing verbal agreement inflection in the computerised picture-choice task

VV-z	CV-z	C-s	C-z
Plays Goes Skis	Watches Dances Washes	Drinks Gets Eats	Reads Swims Sings

Experimental-Like Fillers

The task included 12 pairs of experimental-like fillers, six of which were similar to the past tense trials and the other six were similar to the verbal agreement trials. These fillers were experimental-like in the sense that the same auditory stimuli of the experimental trials were used here but with a bare verb instead of the inflected verb. The following is an example of such pair of trials.

(6.3)

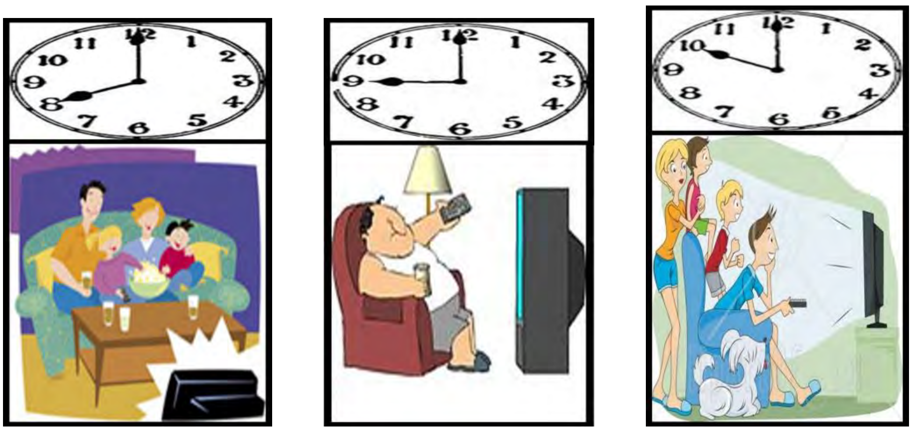
a: Distractor Trial

It’s ten o’clock right now. Listen and choose:

“That happened one hour ago”

b: Experimental-like Filler Trial

“They watch a funny show on TV”



As can be seen here, (6.3) is similar to the trials testing past tense morphology (see example 6.1 above). The difference however is that the main verb lacks the past tense inflection. The inclusion of such trials serves to prevent participants from developing a strategy for responding to the experimental trials, which (if occurred) would give a deceptive indicator of their perception. The strategy feared was that participants would choose the 8 or 9 o'clock pictures every time they are presented with pictures headed with clocks in experimental past tense trials or choose the picture depicting one character doing an action every time they see two pictures depicting the same action by a different number of characters in experimental agreement trials. If such a strategy is developed by participants, it would be observed in the responses to these experimental-like fillers. I did not observe this in the performance of any of the participants.

Training session trials

Five pairs of trials were designed to provide participants with some training to familiarise them with the task. These trials focused on drawing participants' attention to the time displayed in clocks (8, 9 or 10 o'clock) as in (6.4), the number of characters within pictures (s/he vs they) as in (6.5) and the time of the action (now or one/two hours ago) as in (6.6).

(6.4)

a: Trial One

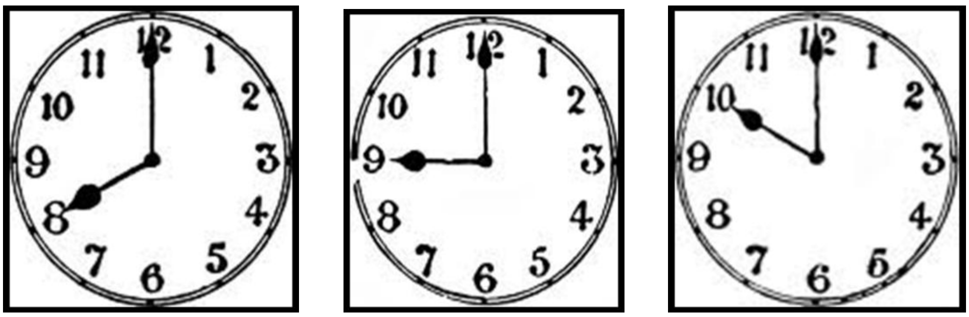
It's ten o'clock right now. Listen and choose one picture:

"What time is it now?"

b: Trial Two

It's ten o'clock right now. Listen and choose one picture:

"What time was it one hour ago?"



(6.5)

a: Trial One

"They are very nice"

b: Trial Two

"She is beautiful"

(6.6)

A: Trial One

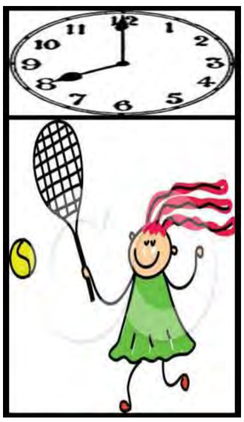
It's ten o'clock right now. Listen and choose one picture:

"She was playing tennis two hours ago"

B: Trial Two

It's ten o'clock right now. Listen and choose one picture:

"She is playing tennis now"



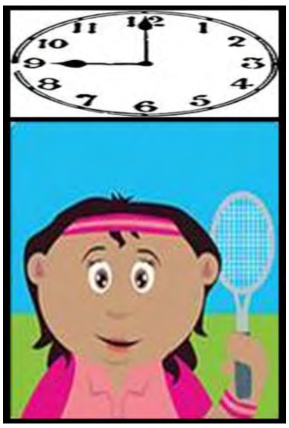


Table 6.3 summarises the types of trials in the computerised picture choice task, their number and purpose.

Table 6.3: Types of trials, their number and purpose in the computerised picture-choice task

Type of trials	Number		Purpose
Training trials	10	5 pairs*	to familiarize participants with the task, drawing participants' attention to time of action and number of characters in pictures
Experimental past tense trials	15	15 pairs*	to test past tense perception in regular and irregular verbs and in verbs creating four different types of coda structure
Distractors to past tense trials	15		to precede experimental past tense trials, distract the participants' attention from the purpose of the task and introduce the tense context
Experimental verbal agreement trials	12	12 pairs*	to test verbal agreement perception in verbs creating four different types of coda structure
Distractors to verbal agreement trials	12		to precede or follow experimental verbal agreement trials, distract the participants' attention from the purpose of the task
Experimental-Like Filler trials	12	12 pairs*	to create trials similar to the experimental ones but with uninflected verbs
Distractors to experimental-like filler trials	12		to precede or follow experimental-like filler trials to make them similar to the distractors to experimental ones.

* pairs share the same visual, but not verbal, stimuli

The test sentences were recorded by native speakers (one female and one male) of British English and then digitised. For clarity reasons and to make it easier for participants to distinguish between the instructions and the stimuli that require their response, the instruction sentence *"it's ten o'clock right now. Listen and choose one picture"* was recorded by a female speaker and the rest of the test stimuli were recorded by a male speaker. All of the visual stimuli used in this task were taken from Google images at http://www.google.co.uk/imghp.

Auditory and visual stimuli presentation: An important part of the task design was the stimuli presentation. My aim was to record participants' behavioral responses on test trials in addition to their reaction times and eye movements (the two latter types of data are presented in Chapter seven). To control the presentation of auditory and visual stimuli, the experiment was designed and run using MATLAB software (The MathWorks Inc. 2008). The design was as follows: At first, the three pictures appeared on screen one after the other in 1.5 seconds. After the three pictures were altogether on screen for 1.5 seconds the auditory stimulus played followed immediately by a beep sound of 200 ms duration to alert participants to give a response. The response was given using the number keyboard buttons 1, 2 and 3, which corresponded to the first, second and third pictures respectively. Test trials were organised in a random but fixed order and it was ensured that no two same-type trials were presented closely after each other. Stimuli presentation is schematically represented as follow:

picture 1.➔	picture 2.➔	picture 3.➔	beep sound➔	auditory stimulus➔	beep sound|| response
↓	↓	↓	↓	↓	↓
1.5 sec	1.5 sec	1.5 sec	200ms	variable length	200ms

6.3.5 Procedure

This task was performed on an individual basis in an eye-tracking laboratory in the neurosciences department at Newcastle University. Participants were seated in a chair facing a 17 inch desktop-computer LCD monitor affixed with a monocular ViewPoint EyeTracker ® (by Arrington Research, Inc.). The eye tracker was also supported with a head-rest positioned in front of the monitor at a distance of 40 cm.

Participants were told that in each trial of the test they would view three pictures and listen to an aural sentence. They were instructed to choose the picture they believed to match the aural sentence or to give their best guess if they were not sure of the correct picture. Test trials were also time locked for 10 seconds. At first, participants completed the training session and then took a break, during which any questions they had were answered. The remaining trials were completed in three sessions interspersed with breaks. This task was completed in no more than 35 minutes in total.

Three types of data resulted from the computerised picture-choice task. These are the behavioral responses, reaction times and eye movements. We present hereafter how the behavioral response data were coded and analyzed leaving the other two types of data to Chapter seven.

Behavioral response: If the target picture was chosen, the answer was coded as correct and if the competitor picture was chosen, the answer was coded as incorrect. The percentage of correct perception was calculated out of the total number of correct and incorrect responses.

There were two occasions where the response to experimental trials was excluded. The first occasion was when the foil picture was chosen. The rationale for this is that choosing the foil picture does not tell anything about the perception of the morphology under investigation. The type of knowledge we are investigating is the perception of past tense and verbal agreement morphology and a pre-requisite for such an investigation is the comprehension of the aural stimulus as a whole. Since the pictures presented in each trial contained a foil picture alongside the two competing pictures (i.e., target and competitor), choosing the foil picture means a failure in comprehending the aural stimuli and not a failure in perceiving the morphological paradigm on the verb.

The second occasion where the response was excluded was the response to an experimental past tense trial when the distractor trial preceding it received an incorrect answer. As elaborated in 6.3.4, experimental past tense trials were preceded by distractor trials in which the context of past tense was introduced as "*it's ten o'clock right now. Listen and choose: X was doing something one hour ago/two hours ago*." An incorrect answer to such trials indicates that the participant was not aware of the time context and, accordingly, the answer to the next experimental trial was not reliable whether it was correct or not. Therefore, it was excluded.

Two Arab participants dropped out before completing the task, so their data from this task were excluded. Data from three participants (1 Arab and 2 Chinese) were also excluded due to a large number of observations of responses as described in the two occasions above. The data from only the past tense trials were excluded for one participant (Chinese) for the same reason. For the rest of the participants, only 21 out of total number of 1755 experimental trials were excluded in line with the same procedure.

The following section presents the results of this study.

6.4 Results

This section reports on the results from the behavioural response to the computerised picture-choice task testing the perception of past tense and verbal agreement morphology by Arab and Chinese speakers at three levels of L2 proficiency. Overall scores, developmental patterns with rising proficiency and L1 effects will be presented. To examine if any phonological constraints affect the perception of the inflection, overall results will be broken down by verb and coda types.

Similar to production data, the perception data were coded into SPSS and tested for non-normal distribution. Again Shapiro-Wilk test revealed non-normal distribution of many variables. Therefore, non-parametric analyses were used for comparisons across variables.

The past tense morphology perception results are presented first.

6.4.1 Past Tense Perception: Arab vs. Chinese

As elaborated in 6.3.4, the computerised picture-choice task contained 15 (out of the total 88) trials that had verbs inflected for past tense. Table 6.4 presents the accuracy scores and percentages for all non-native groups and the control group.

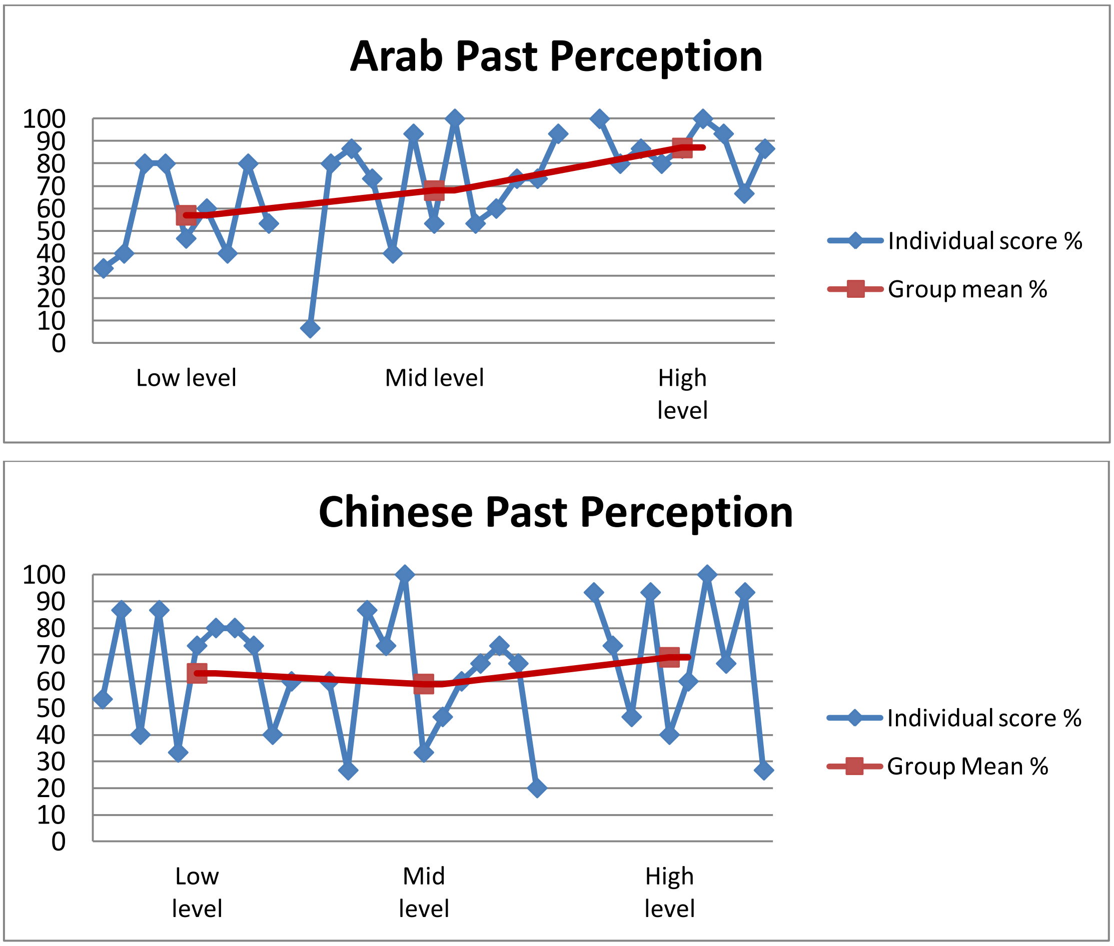
Table 6.4: Scores, percentages and standard deviations of past perception by all groups (number of participants in each group in parenthesis)

Arabic				Chinese				Native (n=10)		
	Score	%	SD		Score	%	SD	Score	%	SD
Low (n=9)	77/135	57.03	19	Low (n=11)	102/161	63.35	20	139/150	92.66	10
Mid (n=13)	133/195	68.20	26	Mid (n=13)	107/180	59.44	24			
High (n=9)	117/135	86.66	10	High(n=10)	104/150	69.33	26			

Table 6.4 shows that the control group perceived the morphology accurately as expected. It also demonstrates that the perception of morphology was highly variable by Arab and Chinese groups at the Low and Mid proficiency levels and while the Arab High-level participants perceived the inflection accurately, their Chinese counterparts exhibited variability.

On a par with production results, great SDs are observed in the perception results as shown in Table 6.4, indicating wider distances between the group mean and the individual results. This can be particularly observed in the results of Chinese participants at all levels and the Arab participants at Low and Mid levels. To gain a better understanding of these results, individual scores along with trajectories of development across group means are illustrated in Figure 6.4.

Figure 6.1: Past tense perception by non-natives: Individual scores and group means



Great individual variation can be seen here. This is consistent with these participants' performance in the production of morphology. This will be discussed further in Chapter eight in light of the findings of production, perception and processing studies. Despite this variation, however, a general trend can be seen in Figure 6.1. That is, for Arab participants, there is an increase in the perception of past morphology with rising proficiency. For Chinese participants, by contrast, no development can be observed. To determine if this is statistically confirmed, Kruskal-Wallis tests were performed on Arab and Chinese participants' results separately. Results from the Arab groups showed the following:

- There was a significant difference among Arab proficiency groups in the perception of the past tense morphology (χ^2=9.183, p<.05, η^2=.30). Post hoc testing (Mann-Whitney) indicated that the Low and Mid-level groups were not statistically different (*U*=38, p=.170), the Low-level group performed significantly below the High-level group (*U*=6.5, p<.005, r=.71) and the Mid level group performed at an approaching significance level below the High-level group (*U*=31, p=.065, r=.39).

Results from the Chinese groups showed the following:

- There was no significant difference among Chinese proficiency groups in the perception of past tense morphology (χ^2=.971, p=.615).

That is, only the Arab participants' perception of past developed with rising proficiency.

To examine whether there was L1 effect on the perception of past tense morphology, a comparison between Arab and Chinese participants at the same proficiency level was performed. Figure 6.2 is a chart representing this comparison based on the results presented in Table 6.4. The native speakers' results are also included.

Figure 6.2: Past tense perception: Arab, Chinese and native groups' comparison

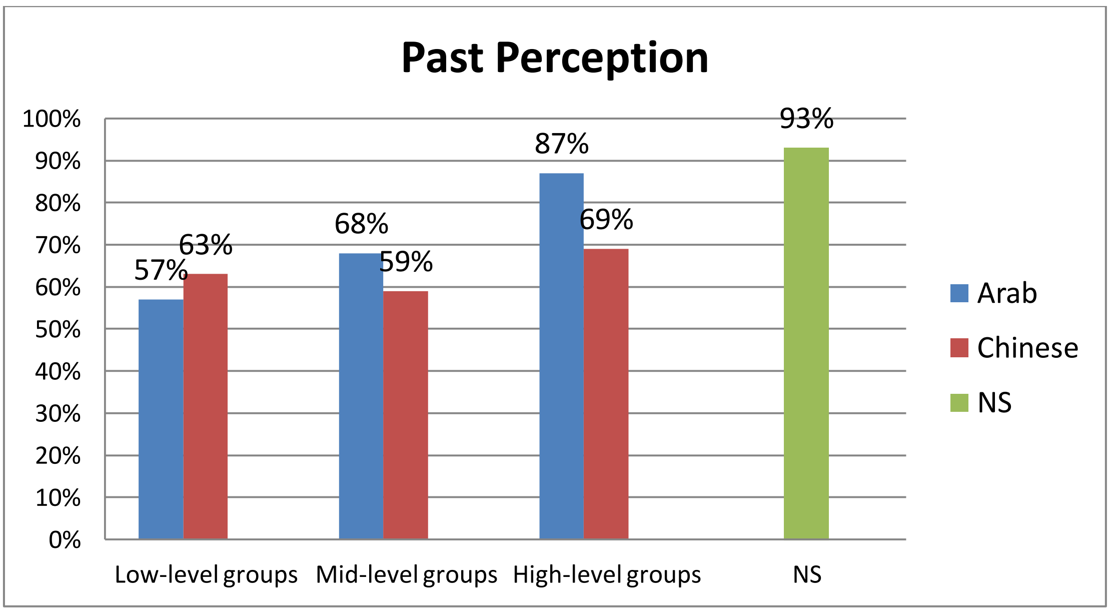


Figure 6.2 shows that there were some differences between Arab and Chinese participants at all proficiency levels. To determine the significance of these differences, Mann-Whitney tests comparing Arab and Chinese groups at the three proficiency levels were performed.

The results showed the following:

- There was no significant difference between the Arab and Chinese groups at any level (Low level: U=39, p=.446; Mid level: U=59, p=.301; High level: U=29, p=.202).

Therefore, no statistical difference between the two language groups at any level was detected. However, a descriptive difference between the Arab and Chinese groups at the high level is noticeable (87% vs 69%, respectively). This was not confirmed statistically probably due to the wide variation of individual results within the Chinese group (SD = 26%) in comparison to the smaller variation within the Arab group (SD = 10%). One way to understand how both groups differed from each other is checking how they performed relative to native speakers. To this end, using Mann-Whitney test, all groups were separately compared to native speakers. The results showed the following:

- There was no significant difference between the Arab High-level group and the control group (U=27, p=.129), but there were significant differences between all other groups and the control group (Arab Low vs. NS: U=6, p<.005, r=.74; Arab Mid vs. NS: U=22, p<.01, r=56; Chinese Low vs. NS: U=11, p<.005, r=.68; Chinese Mid vs. NS: U=13, p<.005, r=.67; Chinese High vs. NS: U=17, p<.05, r=.56).

This indicates that only Arab participants at the high level performed in a native-like manner. The following table (Table 6.5) shows how many participants in each group performed in the native speakers' range. The native speakers' results ranged between 80% and 100% in response to past tense items.

Table 6.5: Individual performance within and below NS range in past perception

	Arab Low		Ch Low		Arab Mid		Ch Mid		Arab High		Ch High	
	N	%	N	%	N	%	N	%	N	%	N	%
Within NS range	3/9	33	4/11	36	5/13	38	2/12	16	8/9	89	4/10	40
Below NS range	6/9	67	7/11	64	8/13	62	10/12	84	1/9	11	6/10	60

(Ch = Chinese; N = number of participants out of the total number of participants in the same group)

Table 6.5 shows that 8 out of the 9 Arab participants (89%) at the High level performed within the native speaker's range. By contrast, only 4 out of the 10 Chinese participants (40%) at the same level performed within the native speakers' range.

To sum up, the results of the computerised picture-choice task showed that in the perception of past, while Chinese participants experienced perceptual difficulty at the Low, Mid and High levels of proficiency, Arab participants did so only at the Low and Mid levels, overcoming the difficulty at the High level. It was also found that for Chinese participants no development in the perception of past occurred with rising proficiency whereas for Arab participants remarkable progress took place between the Low and High levels. The comparison between Arab and Chinese participants showed that there were no statistically significant differences between them at any proficiency level; however, only the Arab High level learners were similar to native speakers with their Chinese counterparts performing significantly below.

The results of the verbal agreement perception are presented below.

6.4.2 Verbal Agreement Perception: Arab vs. Chinese

As described in 6.3.4, the computerised picture-choice task contained 12 (out of the total 88) trials that had verbs inflected with the third-person agreement *–s*. Table 6.6 reports the accuracy scores and percentages for all non-native groups and the control group.

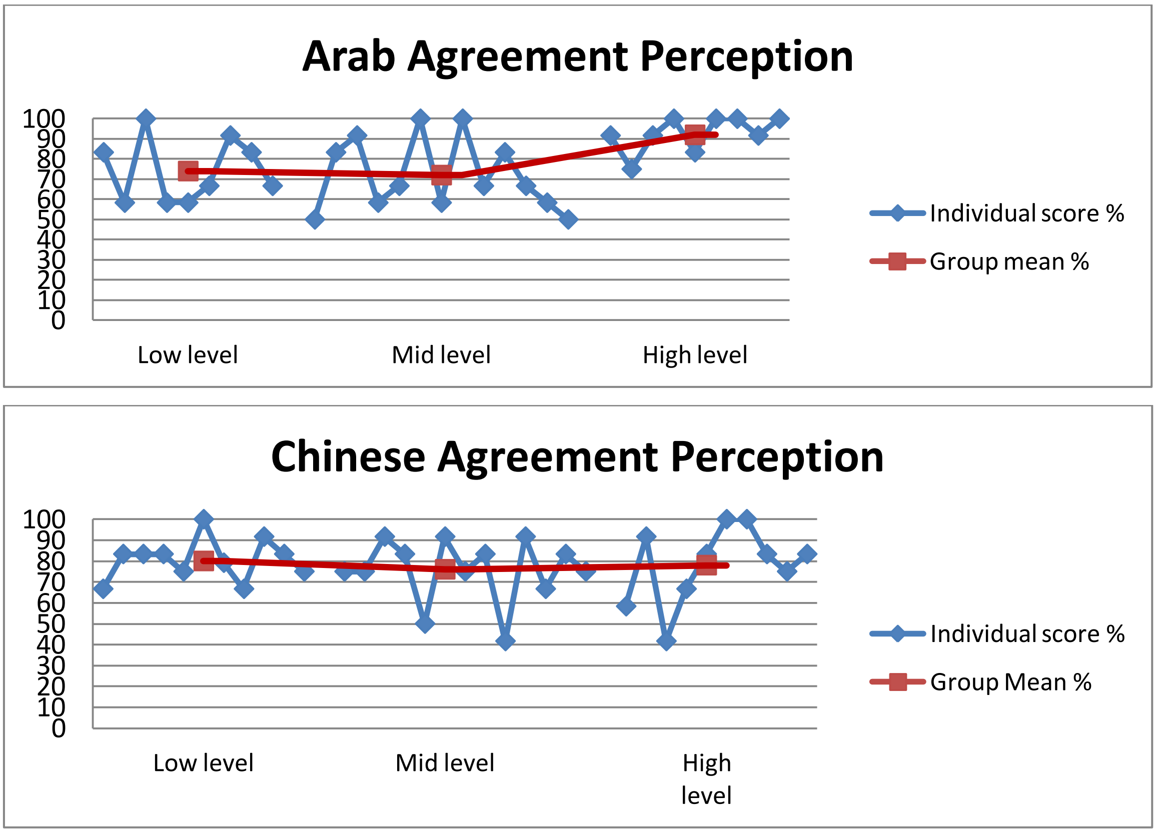
Table 6.6: Scores, percentages & standard deviations of agreement perception by all groups (number of participants in parentheses)

Arab				Chinese				Native (n=10)		
	Score	%	SD		Score	%	SD	Score	%	SD
Low (n=9)	80/108	74.07	16	Low (n=11)	104/130	80	10	117/120	97.5	4
Mid (n=13)	112/156	71.79	18	Mid (n=13)	118/156	75.64	15			
High (n=9)	100/108	92.59	9	High (n=10)	94/120	78.33	18			

The control group performed at ceiling level as expected. The non-native groups at the Low and Mid proficiency levels exhibited perceptual difficulty. However, at the High level, while the Arab group performed accurately, its Chinese counterpart experienced difficulty in perceiving the inflection. These results show similar patterns to those attested in past perception. The agreement perception is higher overall though

Again, large SDs are noticed in many group results as shown in Table 6.6. Figure 6.3 depicts individual results and visualizes the developmental trajectories with rising proficiency for Arab and Chinese participants.

Figure 6.3: Agreement perception by non-natives: Individual scores and group means



The developmental trajectory of Chinese groups presented in Figure 6.3 shows no progress over the three proficiency levels. By contrast, the trajectory of Arab groups shows a different trend; while no development is observed to have occurred between the Low to Mid levels, a remarkable improvement takes place at the High level. To determine the significance of the development in the perception of the inflection across different proficiency groups, Kruskal-Wallis tests were performed on Arab and Chinese group results separately. Results from the Arab groups showed the following:

- There was a significant difference among Arab groups in the perception of the verbal agreement inflection ($\chi^2(2)=8.274$, $p<.05$, $\eta^2=.27$). Post hoc testing (Mann-Whitney) indicated that the Low and Mid-level groups were not statistically different ($U=53$, $p=.710$) but both performed significantly below the High-level group (Low vs. High: $U=13$, $p<.05$, $r=.57$; Mid vs. High: $U=20$, $p<.05$, $r=.54$).

Results from the Chinese groups showed the following:

- There was no significant difference among Chinese groups in the perception of verbal agreement inflection ($\chi^2(2)=.358$, $p=.836$).

Therefore, development in the perception of third-person agreement *–s,* similar to that of past tense morphology, appeared in the performance of only Arab participants with rising proficiency.

To examine whether there was an L1 effect on the perception of the third-person *–s*, a comparison between Arab and Chinese learners at the same proficiency level was performed. Figure 6.4 is a chart representing this comparison based on the results presented in Table 6.6. The native speakers' results' are included.

Figure 6.4: Agreement Perception: Arab, Chinese and control groups' comparison

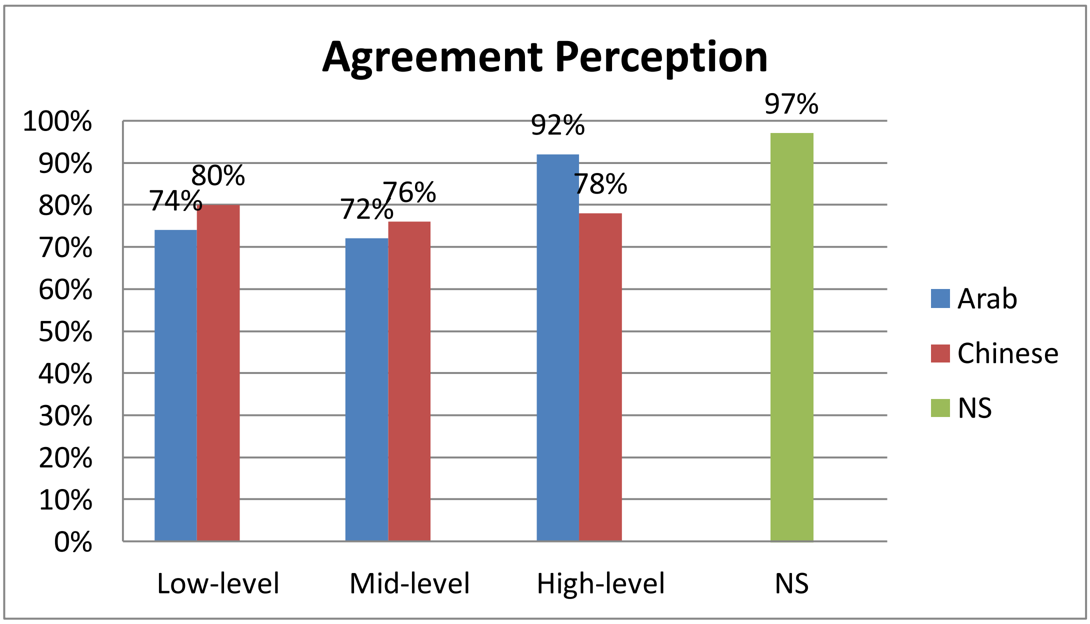


Figure 6.4 shows little differences between Arab and Chinese Low-level and Mid-level groups. At the High proficiency level, however, the Arab group performed better than its Chinese counterpart. Moreover, the only performance that is approaching the control group's performance is that of the Arab High-level group. To determine significance, Mann-Whitney tests comparing Arab and Chinese groups at the three proficiency levels were performed.

- There was no significant difference between the Arab and Chinese learners at the Low level (U=35, p=.263) or the Mid level (U=72, p=.535), but there was an approaching-significance difference between the High-level groups (U=22, p=.060, r=.43).

To check how these participants performed in comparison to native speakers, Mann-Whitney tests were used to compare all groups separately with the control group. The results showed the following:

- There was no significant difference between the Arab High-level group and the control group (U=29, p=.129), but there was a significant difference between each of all other groups and the control group (Arab Low vs. NS: U=8, p<.005, r=.72; Arab Mid vs. NS: U=14, p<.005, r=.68; Chinese Low vs. NS: U=9, p<.005, r=.73; Chinese Mid vs. NS: U=6, p<.001, r=.78; Chinese High vs. NS: U=14, p<.005, r=.64)..

This is the same pattern attested in the past tense results. Only Arab participants at the High level performed in a native-like manner. To further understand how these groups were similar or different from native speakers, Table 6.7 shows the number of participants in each group who performed within or below the native speaker's range. The NS results ranged from 91.67% to 100%.

Table 6.7: Individual performance within and below NS range in verbal agreement perception

	Arab Low		Ch Low		Arab Mid		Ch Mid		Arab High		Ch High	
	N	%	N	%	N	%	N	%	N	%	N	%
Within NS range	2/9	22	2/11	18	3/13	23	3/13	23	7/9	78	3/10	30
Below NS range	7/9	78	9/11	82	10/13	77	10/13	77	2/9	22	7/10	70

(Ch = Chinese; N = number of participants out of the total number of participants in the same group)

It can be seen here that at the High level while 7 out of the 9 Arab participants (78%) performed within the NS range, only 3 out of the 10 Chinese participants (30%) did so.

The perception of verbal agreement morphology showed similar patterns to those attested in the perception of past tense morphology; while Chinese participants experienced perceptual difficulty at the Low, Mid and High levels of proficiency, Arab participants did so only at the Low and Mid levels, overcoming the difficulty at the High level. It was also found that for Chinese participants no development in the perception of verbal agreement occurred with rising proficiency, whereas for Arab participants remarkable progress took place between the Low and Mid levels on the one hand and the High level on the other. The comparison between Arab and Chinese groups showed that there were no statistically significant differences between the Low or Mid level groups but there was a trend towards a significant difference between the two High level groups, with the Arab group performing better. Also, only the Arab High level participants were similar to native speakers with their Chinese counterparts performing significantly below.

Now, I turn to check if the perceptual difficulty and/or the difference observed between the Arab and Chinese groups can be attributed to phonological decoding problems affecting specific shape types of the coda. To this end, the results will be broken down by allomorph type, cluster type and (in case of past tense) verb (ir)regularity.

6.4.3 Past Perception: Allomorph Type and Simple vs. Complex Codas

This section presents the results on past tense perception broken down by the four allomorphs of the *–ed* inflection and consonant cluster vs. non-cluster endings. Table 6.8 reports the Arab groups' results

Table 6.8: Perception of past tense *–ed* by allomorph among Arab groups.

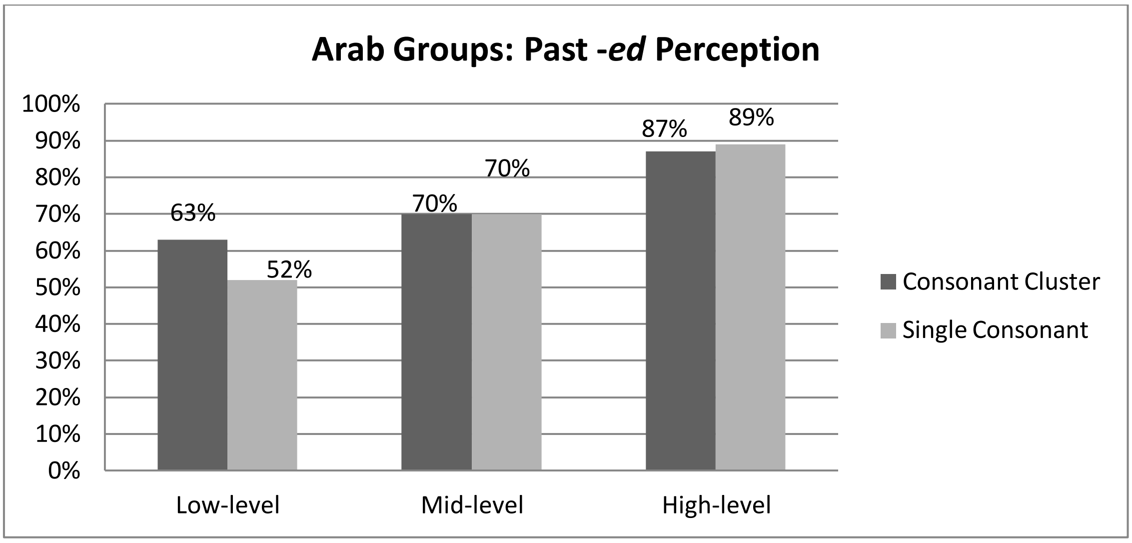
	CVV-d		CV-d		C-d		C-t	
Arabic	Score	%	Score	%	Score	%	Score	%
Low (n=9)	13/27	48.14	15/27	55.55	17/27	62.96	17/27	62.96
Mid (n=13)	27/39	69.23	28/39	71.79	28/39	71.79	27/39	69.23
High (n=9)	23/27	85.18	25/27	92.59	25/27	92.59	22/27	81.48

The phonological form of the verb seems to have little effect on the perception of the past tense inflection as Table 6.8 shows small differences among the rates of perception across the four allomorphic variants of *–ed* for the three Arab groups. To test for significance, Friedman tests were carried out. The results showed the following.

- There were no significant differences in the perception of the past tense *–ed* across its four allomorphic variants for Arab participants at any level: Low level ($\chi^2(3)=1.095$, p=.778), Mid level ($\chi^2(3)=.176$, p=.981) or High level ($\chi^2(3)=3.267$, p=.352).

The scores of consonant cluster forms (C-d & C-t) and non-cluster forms (VV-d & CV-d) were combined and Figure 6.5 represents the Arab groups' results.

Figure 6.5: The perception of past tense –ed in word-final consonant cluster and non-consonant cluster contexts by the three Arab groups.



To determine whether there was any significant differences in the results, Wilcoxon signed rank tests were performed. Results were as follows:

- There were no significant differences in the perception of the past tense inflection between cluster and non-cluster contexts for Arab learners at any level: Low level (Z=-1.409, p=.159), Mid level (Z=-.274, p=.784), or High level (Z=-.632, p=.527).

The Chinese participants' results of the perception of the past tense *–ed* inflection across its four allomorphs are presented in Table 6.9.

Table 6.9: Perception of past tense *–ed* by allomorph among Chinese groups.

	CVV-d		CV-d		C-d		C-t	
Chinese	Score	%	Score	%	Score	%	Score	%
Low (n=11)	15/33	45.45	23/31	74.19	22/32	68.75	20/32	62.5
Mid (n=13)	18/36	50	18/36	50	20/36	55.55	26/36	72.22
High (n=10)	18/30	60	20/30	66.66	22/30	73.33	21/30	70

The phonological form of the verb seems to have some effect on the perception of the past tense inflection as Table 6.9 shows small differences among the rates of perception across the four allomorphs of *–ed* for the three Chinese groups. To test for significance, Friedman tests were carried out. The results showed the following.

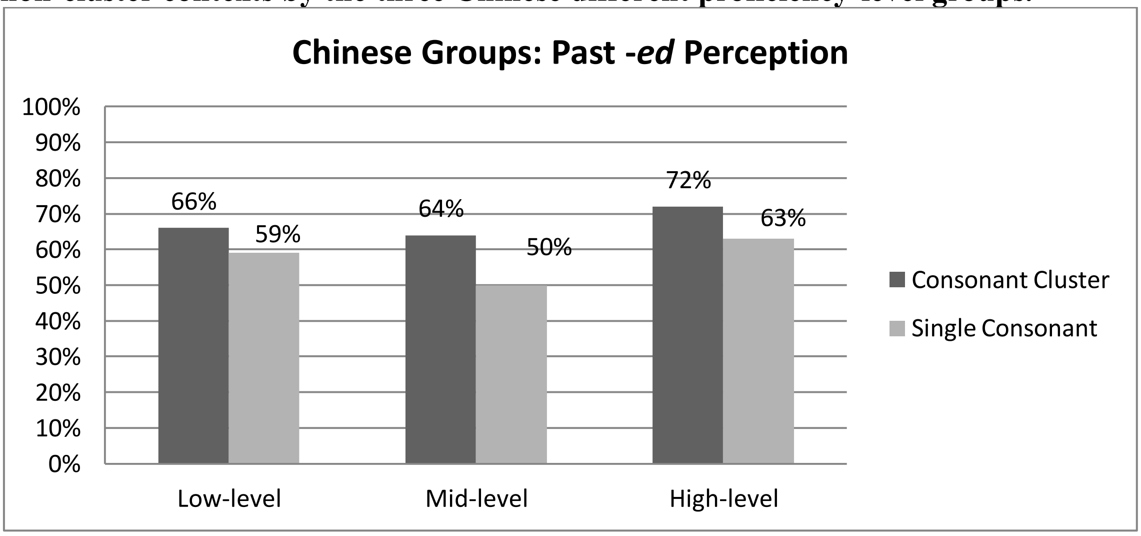
- There was a significant difference in the perception of the past tense *–ed* across the four allomorphs for the Chinese learners at the Low level ($\chi^2(3)=8.103$, $p<.05$, $\eta^2=.18$). Post hoc Wilcoxon signed rank tests showed that the perception of the CVV-d was significantly lower than that of the CV-d ($Z=-2.058$, $p<.05$, $r=.43$), but there were no significant differences between any of the other combinations.

- There was a significant difference in the perception of the past tense *–ed* across the four allomorphs for the Chinese participants at the Mid level ($\chi^2(3)=8.735$, $p<.05$, $\eta^2=.17$). Post hoc Wilcoxon signed rank tests showed that the perception of the C-t allomorph was significantly higher than the perception of CVV-d ($Z=-2.203$, $p<.5$, $r=.43$), CV-d ($Z=-2.203$, $p<.5$, $r=.43$) and C-d ($Z=-2.060$, $p<.5$, $r=.40$).

- There was no significant difference in the perception of the past tense *–ed* across the four allomorphs for Chinese learners at the High level ($\chi^2(3)=1.966$, $p=.580$).

Figure 6.6 represents the Chinese groups' results where the scores of consonant cluster (C-d & C-t) and non-cluster forms (CVV-d & CV-d) are combined.

Figure 6.6: The perception of past tense *-ed* in word-final consonant cluster and non-cluster contexts by the three Chinese different proficiency-level groups.



To determine whether there is a significant difference in the perception of the inflection in consonant clusters and non-clusters, Wilcoxon signed rank tests were performed. Results were as follows:

- There were no significant differences in the perception of the inflection between cluster and non-cluster endings for Chinese participants at any level (Low: Z=-.758, p=.448; Mid: Z=-1.604, p=.106; High: Z=-.768, p=.443)
-

All in all, we can conclude that the perception of past tense *–ed* is very slightly affected by the phonological form of the verb. For Arab participants, the perception was similar for the four allomorphic variants of the *–ed* inflection. For the Chinese participants, it was found that the perception of the inflection was lower in verbs ending with a VV-d structure than other verb forms at the Low level, and perception was higher for verbs ending in C-t than other verb forms at the Mid level. But perception was not different across the four allomorphs at the High proficiency level. Above all, the perception was found to be similar in verbs with consonant cluster and non-cluster endings for both Arab and Chinese groups at the three proficiency levels. These results, therefore, did not indicate that phonological constraints on the perception of word-final consonant clusters could explain the perceptual difficulty experienced by our participants.

I turn now to the verbal agreement perception results broken down by allomorph and cluster types.

6.4.4 Agreement Perception: Allomorph Type and Simple vs. Complex Codas

As described before, the computerised picture-choice task included 12 trials testing for verbal agreement, which were divided into 3 trials for each of the allomorphic variants of the *–s* inflection. Table 6.10 reports the Arab groups' results

Table 6.10: Perception of verbal agreement –s by allomorph among Arab groups.

	CVV-z		CV-z		C-z		C-s	
Arabic	Score	%	Score	%	Score	%	Score	%
Low (n=9)	19/27	70.37	19/27	70.37	23/27	85.18	19/27	70.37
Mid (n=13)	29/39	74.35	29/39	74.35	27/39	69.23	27/39	69.23
High (n=9)	27/27	100	25/27	92.59	25/27	92.59	23/27	85.18

The phonological form of the verb seems to have very little effect on the perception of the third-person *–s* as Table 6.10 shows small differences among the rates of perception across the four allomorphic variants for the three Arab groups. To test for significance, Friedman tests were carried out. The results showed the following.

- There were no significant differences in the perception of the third-person *–s* across its four allomorphic variants for Arab participants at any level: Low level ($\chi^2(3)=2.115$, p=.549), Mid level ($\chi^2(3)=.523$, p=.914) or High level ($\chi^2(3)=3.686$, p=.297).

To compare the perception rates of the inflection between consonant cluster and non-cluster endings, the scores of C-z & C-s forms in one hand and VV-z & CV-z forms on the other were combined. Figure 6.7 gives the overall pattern of the *–s* perception broken down by consonant cluster and non-cluster endings in the performance of the Arab participants.

Figure 6.7: Perception of verbal agreement –s in word-final consonant cluster and non-cluster contexts by the three Arab different proficiency-level groups.

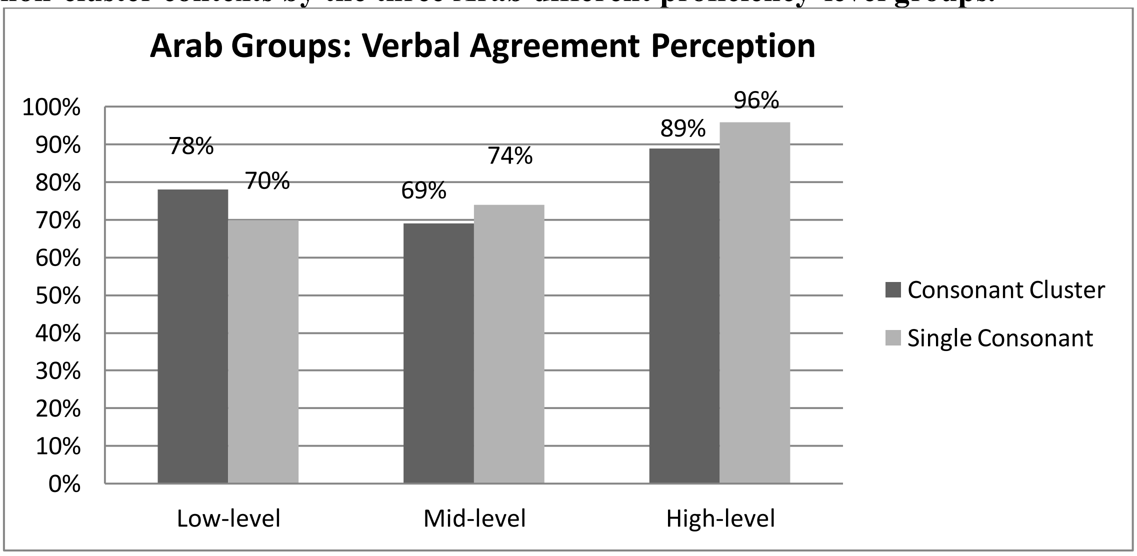


Figure 6.7 shows that there are some small differences between the perception of the inflection in consonant cluster and non-cluster endings in the performance of the three groups. To test for significance, Wilcoxon signed rank tests were performed. Results were as follows:

- There were no significant differences in the perception of the inflection between verbs ending with a consonant cluster and non-cluster for Arab learners at any level: Low level (Z=-.935, p=.350), Mid level (Z=-.493, p=.622) or High level (Z=-1.511, p=.131)

The Chinese participants' results of the perception of the verbal agreement *–s* in its four allomorphic variants are reported in Table 6.11.

Table 6.11: Perception of verbal agreement *–s* by allomorph among Chinese groups.

	CVV-z		CV-z		C-z		C-s	
Chinese	Score	%	Score	%	Score	%	Score	%
Low (n=11)	31/33	93.93	21/32	65.62	26/33	78.78	26/32	81.25
Mid (n=13)	34/39	87.17	26/39	66.66	30/39	76.92	28/39	71.79
High (n=10)	26/30	86.66	20/30	66.66	25/30	83.33	23/30	76.66

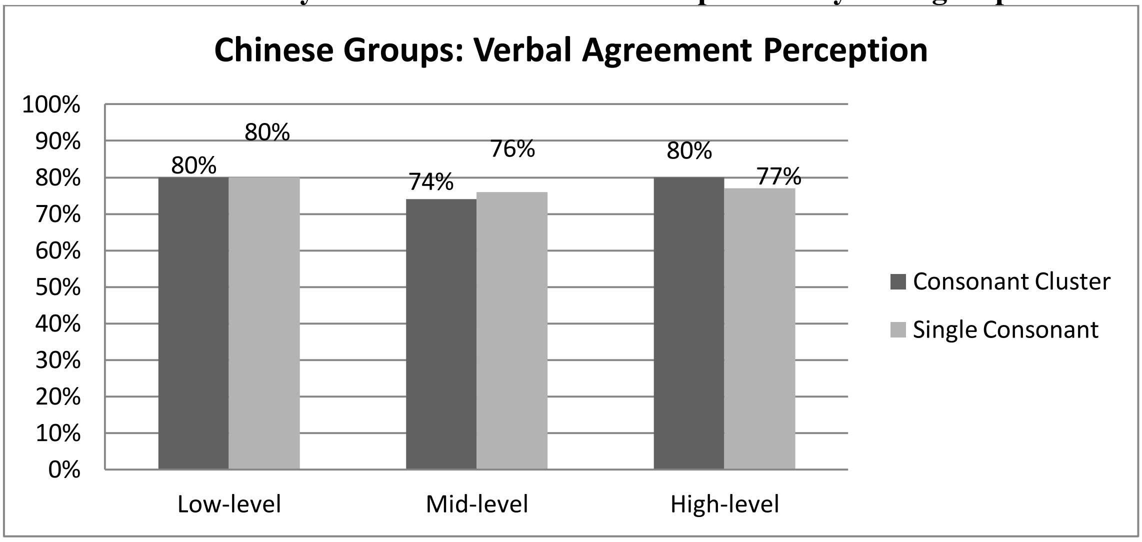
To test for differences, Friedman tests were carried out on the rates of the four allomorphs for each proficiency group separately. The results showed the following.

- There was a significant difference in the perception of the third-person *–s* across its four allomorphic variants in the performance of the Chinese Low-level group ($\chi^2(3)=13.817$, $p<.01$, $\eta^2=.31$). Post hoc Wilcoxon signed rank tests showed that the perception of the VV-z allomorph was significantly higher than its perception in all other allomorphs (CVV-z vs. CV-z: $Z=-2.588$, $p<.05$, $r=.55$; CVV-z vs. C-s: $Z=-2.236$, $p<.05$, $r=.47$; CVV-z vs. C-z: $Z=-2.449$, $p<.05$, $r=.52$) and the perception of CV-z was lower than the perception of C-z and C-s, significantly in the former ($Z=-2.041$, $p<.05$, $r=.43$) and at the approaching significance level in the latter ($Z=-1.802$, $p=.072$, $r=.38$).

- There were no significant differences in the perception of the third-person *–s* across its four allomorphic variants in the performance of the Chinese learners at the Mid level ($\chi^2(3)= 6.479$, $p=.091$) or High level ($\chi^2(3)= 5.224$, $p=.156$).

Again, to compare the perception rates of the inflection between consonant cluster and non-cluster endings, the scores of C-z & C-s forms in one hand and VV-z & CV-z forms on the other were combined. Figure 6.8 gives the overall pattern of the *–s* perception in consonant cluster and non-cluster endings for the three Chinese groups.

Figure 6.8: Perception of verbal agreement –*s* in word-final consonant cluster and non-cluster contexts by the three Chinese different proficiency-level groups.



Very small differences are observed in Figure 6.8 between the perception of the inflection in consonant cluster and non-cluster endings across the three groups. To determine if any statistical differences exist, Wilcoxon signed rank tests were performed. Results were as follows:

- There were no significant differences in the perception of the inflection in verbs ending with consonant clusters and non-clusters for the Chinese learners at any level (Low: Z=-.135, p=.892; Mid: Z=-.586, p=.558; High: Z=-.530, p=.596).

Overall, the perception of third-person –*s* does not appear to be affected by the phonological form of the verb. For Arabic speakers, the perception was similar for the four allomorphic variants of the inflection. For the Chinese speakers, it was found that the perception of the inflection was higher in verbs ending with VV-z structures at the Low level but it was not different across the four allomorphic variants at the Mid and High levels. Moreover, no difference was detected between the perception in verbs ending with consonant clusters and non-clusters in the performance of any group.

These results indicate that the perceptual difficulty detected in the performance of Chinese participants at the Low, Mid and High levels and Arab participants at the Low and Mid levels was not due to any phonological constraints on the shape of the coda, where the inflection is located.

Nevertheless, as in production, I claim that such results do not exclude the possibility of a general phonological effect on word-final consonants, investigated through a comparison between the perception of regular and irregular past tense morphology. The perception rates in regular and irregular verbs are compared below.

6.4.5 *Past Perception: Regular vs. Irregular*

As described in 6.3.4, the computerised picture-choice task contained 15 trials testing past tense out of which 3 tested irregular verbs and the rest tested regular ones. Table 6.10 reports the results of the Arab groups

Table 6.12: Past perception: regular vs. irregular by Arab groups

	Regular			Irregular		
	Score	%	SD	Score	%	SD
Arabic Low (n=9)	62/108	57.40	21	15/27	55.55	17
Arabic Mid (n=13)	110/156	70.51	26	23/39	58.97	36
Arabic High (n=9)	95/108	87.96	13	22/27	81.48	24

Table 6.12 shows that the perception of the regular and irregular past morphology is similar at the Low level, but the perception of the regular morphology is higher than the irregular morphology at the Mid and High levels. To determine whether there is any significant differences between these results, Wilcoxon signed rank tests. Results were as follows:

- There were no significant differences in the perception of past morphology between regular and irregular verbs for Arab learners at any level (Low: Z=-.416, p=.677; Mid: Z=-1.181, p=.238; High: Z=-.602, p=.547).

The Chinese groups' results are presented in Table 6.13.

Table 6.13: Past perception: regular vs. irregular by Chinese groups

	Regular			Irregular		
	Score	%	SD	Score	%	SD
Chinese Low (n=11)	80/128	63.5	26	22/33	66.66	21
Chinese Mid (n=13)	82/144	56.94	24	25/36	69.44	30
Chinese High (n=10)	81/120	67.5	31	23/30	76.66	22

Table 6.13 shows that the perception of the irregular morphology is higher than the perception of the regular morphology at all proficiency levels. To determine whether there is any significant differences between these results, Wilcoxon signed rank tests. Results were as follows:

- For the Chinese learners, the difference in the perception of past morphology between regular and irregular verbs was not significant at the Low level (*Z*=-.356, p=.722), approaching-significance at the Mid level (*Z*=-1.887, p=.059, r=.37) and not significant at the High level (*Z*=-.831, p=.406).

Therefore, overall, none of the Chinese or Arab groups perceived the irregular morphology significantly better than the regular inflection, implying no phonological constraints on the specific perception of the regular past tense morpheme. The trend towards a significant result attested in the performance of Chinese Mid-level participants in perceiving the irregular morphology better was an exception. Yet, it does not endanger the conclusion that no phonological constraints appear to affect the perception of the regular morpheme.

6.5 Summary and Hypothesis Evaluation

The present experiment has investigated the perception of past tense and verbal agreement morphology by Arab and Chinese L2ers of English at three proficiency levels. We have been particularly interested in examining whether the variability, developmental patterns and L1 effects attested in the production of morphology in Chapter five extend to the perception of the same participants. Assuming that production and perception use the same syntactic representations, five experimental hypotheses (based on accounts of variability), parallel to those set up for the production in the previous chapter, were formulated. In this section, the results from the perception task are summarised first and then the five experimental hypotheses are evaluated against the findings.

The behavioural response to the computerised picture-choice task showed that perceptual difficulty, as in Chapter five for production, appeared at Low, Mid and High proficiency levels in the performance of Chinese participants whereas Arab participants experienced this difficulty only at Low and Mid levels overcoming it at the High level. The developmental patterns revealed remarkable differences between Arab and Chinese learners as improvement in the perception of morphology was observed only in the performance of Arab participants (at the High level of proficiency), while Chinese participants' perception showed no progress with rising proficiency. Comparisons between the two language groups revealed no differences at the Low and Mid levels. For High level participants, no statistical difference was detected in response to past tense items, but the difference approached significance in verbal agreement trials. Other L1 based differences were found in the results of High level groups as only Chinese participants experienced a perceptual difficulty and while Arab participants at the High level performed statistically similarly to native speakers, their Chinese counterparts performed significantly below.

Furthermore, pursuing the source of perceptual difficulty, the results were broken down by 1) allomorphic variants (VV-d/z, CV-d/z, C-t/s and C-d/z), 2) cluster type (consonant cluster and non-cluster) and 3) verb type (regular and irregular). The results showed marginal phonological effects. Results of these analyses showed the following:

1) Allomorphic variants: The perception of all allomorphs of past and agreement inflections was variable by all groups except Arab High level participants. Disparity in their rates was observed in three instances. These are a) VV-d lower than CV-d in the performance of Chinese Low level participants, b) VV-z was higher than CV-z, C-d and C-t and C-z was higher than CV-z in the performance of Chinese Low level participants and c) C-t was higher than VV-d, CV-d and C-d in the performance of Chinese Mid level participants.

2) Cluster type: The perception of past and agreement inflections in cluster and non-cluster contexts was the same in the performance of all groups.

3) Verb type: The perception of past and agreement was the same for regular and irregular verbs in all groups

To tie these results directly to the research hypotheses formulated in section 6.2, they will be evaluated here one by one, and connected to the relevant underlying models and theories. For convenience, the research hypotheses will be repeated below and each will be followed by conclusions.

H1. The perception of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, and both language groups will improve in accuracy similarly with rising proficiency.

This hypothesis (based on OG) is partially supported by the perception results. Only the Arab participants performed as predicted as they perceived the inflection variably at lower proficiency levels and became more accurate with rising proficiency. On the other hand, the Chinese groups' performance constitutes a problem for this hypothesis. Specifically, no improvement in the Chinese participants' perception took place with higher proficiency which is against what the hypothesis predicts.

H2. The perception of morphology will be variable at lower proficiency levels by Arab and Chinese participants alike, but Arab participants will improve faster than their Chinese counterparts with rising proficiency.

This hypothesis (based on the MSBH) is confirmed by these results. As predicted, both Arab and Chinese groups perceived the morphology variably at Low and Mid levels and a significant improvement occurred at the High level with Arab but not Chinese participants.

H3. Arab participants will have an L1-based advantage over their Chinese counterparts in the perception of morphology from lower levels onwards.

This hypothesis (based on the RDH) is also partially supported. An L1 advantage for Arab participants was indeed observed but this took place only at the High level. That Arab and Chinese participants at the Low and Mid levels perceived the morphology similarly variably contradicts this hypothesis.

H4. The perception of morphology will be accurate from early levels onwards by both Arab and Chinese groups alike.

This hypothesis (based on the MSIH) is not supported. The perception of all groups apart from the Arab High level group was variable. Moreover, this variability could not be explained by phonological constraints on complex codas.

> *H5. The perception of the inflection by the Chinese, but not the Arab, participants will be lower in cluster than non-cluster codas (in past tense and agreement contexts) and regular than irregular verbs (in past tense contexts).*

This hypothesis (based on the PhRH) is also not supported. None of the Chinese (or Arab) groups perceived the morphology better in non-cluster codas or irregular verbs than consonant cluster codas or regular verbs, respectively.

In Chapter eight, the results reported in this chapter will be discussed in light of the previous research reviewed in Chapters two and three.

The following chapter presents the processing study, which will show how the morphosyntactic properties under study are processed by the research participants.

Chapter 7. The Processing of Past Tense and Verbal Agreement: The Study Design and Results

7.1 Introduction

The processing data collected from the participants of this research have the power to confirm (or disconfirm) the conclusions drawn from the results reported on in the previous two chapters (Chapters five and six). In Chapter five, we found that all participants, apart from the Arab High level participants, experienced variability in their production of the morphology under study. Chapter six revealed that this variability correlated with a perceptual difficulty. Moreover, the data reported in both chapters showed that production and perception difficulties could not be associated with (L1 transferred) phonological constraints. An L1 effect was observed only at the High level with the Arab participants performing better than their Chinese counterparts. These findings gave more weight to syntactic than non-syntactic accounts of the phenomenon of morphological variability. They particularly indicated that the relevant syntactic representations were more likely to be absent at lower proficiency stages and become instantiated with rising proficiency. Also, the instantiation of these syntactic representations seemed to be influenced by the presence or absence of similar representations in the learners' L1 as this took place faster for Arab than Chinese participants. Based on the assumption that the three modalities, production, perception and processing, use the same syntactic representation, it is conceivable that processing data should demonstrate similar patterns as those observed in production and perception data.

To test the processing of past tense and verbal agreement morphology, the computerised picture choice task described in Chapter six (section 6.3.4) was affixed with reaction time and eye-tracking measures. The behavioural response to the task (choosing a picture) was taken as an indicator of participants' perception of the morphology and this was reported in Chapter six. In this chapter, the data from the reaction time and eye-tracking measures will be taken as an indicator of how participants process the morphology under study.

This chapter is organised as follows. Section 7.2 outlines the methodology of the processing study. The experimental hypotheses are formulated in section 7.3. Next, section 7.4 presents the results of the experiment. This is followed by a summary of the findings by which the experimental hypotheses are evaluated in section 7.5.

7.2 Method

7.2.1 Design

The reaction times (RTs) and eye movements of the participants upon their response to the trials of the computerised picture choice task used in the perception study (Chapter six) were recorded. In this task, participants were presented with sets of three pictures accompanied by aurally presented sentences and they were asked to choose the picture they considered best depicted by the oral stimuli. RTs to test stimuli and eye movements during its presentation until a response is given are assumed to show how participants process the morphology under study revealing information on the nature of the corresponding underlying syntactic representations.

7.2.2 Participants

The participants of this study are the same native and non-native speakers who took part in the perception study presented in Chapters five and six. These were a control group of 10 native speakers of British English and 71 non-native participant, who were native speaker of Arabic (n= 34) or Mandarin Chinese (n= 37). Participants' background information and the non-native proficiency details are presented in Chapter four, sections 4.3.1 and 4.4, respectively. The reader is referred to those sections for more details.

7.2.3 Reaction Time and Eye Movement: Theoretical Assumptions

The present experiment used reaction time (RT) and eye-tracking measures performed on a computerised picture choice task to test the processing of past tense and verbal agreement morphology in SLA of English. The use of these measures in the field of

SLA is relatively new and, therefore, prior to presenting how they are applied on the picture choice task in the present experiment, it is essential to elucidate their theoretical underpinnings and explain how the data from these measures were used to inform our investigation.

RT is a measure of the speed at which a participant responds to test stimuli, and in the field of L2 processing it is assumed to be the time it takes to process the linguistic information "with the idea that speed of response is an indication of processing load" (Gass and Selinker, 2008: 63). It is widely believed that inferences on the L2 mastery of linguistic features can be derived from the RT to stimuli (Ellis and Barkhuizen, 2005). This is based on the assumption that mastering a linguistic feature means developing an automatic competence in it. According to Jiang (2007: 2), automatic competence "refers to the ability to apply one's linguistic knowledge spontaneously in both the productive and receptive use of language". Therefore, in L2 processing, a mastery of a certain feature means developing an ability to make automatic real-time use of that feature in production and comprehension, which would reflect on the speed of the reaction to test stimuli.

Developing automatic competence for a specific feature is the same as building mental representations for it, but viewed from two different perspectives (Jiang, 2007). These perspectives are L2 processing and L2 acquisition; while mastery in L2 processing is viewed as developing automatic competence as discussed above, mastery in L2 acquisition is seen as building mental representations. In Chapters five and six of this thesis, I took consistent production and accurate perception of past tense and verbal agreement morphology as an indicator of the instantiation of target-like mental representations for these properties and, conversely, the inconsistent production and inaccurate perception was seen as a reflection of non-target-like representations. Likewise, in this chapter, quick RTs to stimuli which received correct decision will be taken as an index of the development of automatic competence and slow RTs will be viewed as absence of automatic competence. The decision on whether an RT is quick or slow will be taken based, as is often the case in the field of L2 processing, on the RTs of native speakers performing the same task.

Another method this chapter is concerned with is eye-movements (also called eye-tracking). This can be divided into two methods, which are 1) eye-movements on written text during reading and 2) eye-movements on a visual stimulus (i.e., the visual-world paradigm). Recording eye-movements during reading has been more widely used in L2 research and has provided the field with sharp insights into the nature of linguistic knowledge and how it is put to use in L2 processing (for a recent review on the application of this method in L2 research, see Roberts and Siyanova-Chanturia, 2013). I will, however, focus on the second method 'the eye-movements on a visual stimulus' because it is the method used in the present experiment.

In the field of cognitive science, it is believed that eye movements on a visual stimulus reveal information on the underlying cognitive processes. Research on eye tracking has shown "in a particularly compelling way" that moving eyes to capture visual views "is inextricably linked to the observer's cognitive goals" (Hayhoe and Ballard, 2005: 188). Thus, when presented with a visual scene, individuals move their eyes to focus attention on certain regions of interest and, accordingly, these movements provide insights into the underlying cognitive processes (Liversedge and Findlay, 2000).

In psycholinguistics, eye-movements at visual stimuli as the auditory input unfolds are thought to give information on real-time processing of the language and provide insights into the comprehension of specific linguistic features (Tanenhaus and Trueswell, 2006). Based on this, the eye-movement method has been used to collect moment-to-moment language processing data. In linguistic experiments that track the eye movements of individuals while listening to stories and looking at pictures depicting certain objects, it has been found that participants spontaneously initiate saccades (i.e., the eye movement from one fixation or gaze to another) to pictures immediately after they are named in auditory stories (e.g., Cooper, 1974). For example, in one part of a study conducted by Arnold, Eisenband, Brown-Schmidt and Trueswell (2000), the researchers examined how gender information inherent in the English pronoun influenced its real-time interpretation, giving their participants a picture containing two different-gender characters and auditory stimuli. They found that their participants initiated saccades to the referent within 200 ms after the offset of the pronoun and fixated reliably longer on the target than the competitor referent. This indicates that perceiving a linguistic property works as a trigger for eye movements at the visual stimuli.

Therefore, previous research has made it clear that it is safe to assume that comprehending an auditory stimulus entails initiating saccades to a target corresponding visual stimulus. Based on this, in the present experiment, it will be assumed that perception of the verbal morphology will lead participants to initiate saccades and fixations on the target picture. Based on this (following Arnold et al., 2000), as will be elaborated below, two eye movement measures in the present experiment will be taken. These are (1) the length of eye looks to a target picture (compared to a competitor) and (2) the speed of first look to a target picture. Such data will be taken as an indicator of the nature of the underlying representations and how they are put to use during processing.

7.2.4 Reaction Time and Eye Movement: Materials

To test the processing of past tense and verbal agreement, this study used a computerised picture-choice task affixed with reaction time and eye-movement measures. This is the same task used to test the perception of the research participants. In Chapter six, only the behavioral response (picture choosing) to task trials was reported. In this chapter, reaction times and eye movement will be reported. I will briefly review the task here and the reader is referred to Chapter six (section 6.3.4) for more details on the task.

As elaborated in section 6.3.4, the computerised picture choice task consisted of 88 trials. Each trial administered a set of three pictures and an auditory stimulus. The pictures were (a) a target picture (correct choice), (b) a competitor picture (possible but wrong choice) and (c) a foil picture (unrelated picture). Participants were asked to choose one picture, the choice of which depended on their perception of the verbal morphology under study.

The test trials were divided as follows: (1) 10 training trials, (2) 15 trials testing past tense (3) 12 trials testing verbal agreement (4) 12 experimental-like filler trials and (5) 39 distractor trials. Each of the distractor trials formed a pair with one of the trials in 2, 3, or 4 as it used the same pictures but with a different verbal stimulus.

In past tense trials, each of the three pictures was attached with a clock showing 8, 9, and 10 o'clock respectively. The target and competitor pictures in each trial depicted the same action. The only difference between both was the time shown in the clock attached to each of them. The context of time of each trial was introduced in the distractor trials preceding them.

In verbal agreement trials, the target and competitor pictures depicted the same action (e.g., dancing) but with a different number of characters. Then, in the verbal stimuli, subject-drop sentences and verbs inflected for third-person agreement were used to examine whether participants could recover the subject number depending solely on the inflection attached to the verb.

The experimental-like filler trials were similar to past tense and verbal agreement trials but with one difference only. The difference was that the verb was in its bare form not inflected for past or agreement. The inclusion of such trials was essential to ensure no participant followed a strategy in responding to the experimental trials. Such a strategy could be assuming that all verbs are inflected for past or agreement and thus the best choice would be the picture showing past time (in past time trials) or one character (in agreement trials). Including experimental like fillers with non-inflected verbs would prevent participants from adopting such a strategy and the response to these trials would reveal if any participants had done this.

The presentation of trials was controlled by MATLAB software (The MathWorks Inc. 2008). The three pictures were presented one after the other in 1.5 seconds followed by the verbal stimuli. After that a beep sound of 200ms played to alert the participant to give a response by pressing the 1, 2, or 3 keyboard buttons corresponding to pictures 1, 2 and 3 respectively.

The reaction time was the time between the beep sound and the key press by a participant. This was recorded by the MATLAB software for each trial.

Eye movements were recorded by a monocular ViewPoint EyeTracker ® (by Arrington Research, Inc.) affixed to the 17 inch desktop-computer LCD monitor on which participants performed the task. The eye tracker was supported with a head-rest positioned in front of the monitor at a distance of 40 cm. This took place in an eye-tracking laboratory in the neurosciences department at Newcastle University.

Eye movements were recorded every millisecond from the beginning of stimuli presentation until a response by the participant was provided. The 88 trials were completed by no more than 35 minutes interspersed by three breaks. The eye tracker was calibrated once at the beginning of the session and then the calibration was re-checked after each break.

7.2.5 Procedure

Three types of data resulted from the computerised picture-choice task. These were (1) picture choice response, (2) reaction time and (3) eye movements. The first type was dealt with in Chapter six. I present hereafter how the reaction time and eye movement data were coded and analysed.

Reaction Times: This measured the time participants took to respond to experimental trials. Only reaction times to the experimental trials that received correct answers were included.

Eye movements: Participants' eye movements were recorded at a rate of 220 frames per second from the time the visual stimuli were presented to the time the answer was provided. Visual stimuli were divided into three categories: (1) the target picture: the correct choice, (2) the competitor picture: the picture that is similar to the target with one difference in either the number of characters or the time of the action, (3) foil picture: the unrelated picture. As in Arnold et al.'s (2000) study, two measures are reported here as follows:

i. The total length of time spent on a region of interest: regions of interest were the target, competitor and foil pictures. I calculated the percentage of the time spent looking at the target picture out of the total period of time spent looking at the target and competitor from the time the stimulus was presented to the time the response was given. The stimulus, here, means the verbal stimuli as a whole.

ii. The sensitivity measure: Sensitivity is defined as the time it takes a participant to initiate the first look to the target picture after the aural stimulus is presented. The stimulus here means the verb bearing the inflection. It is attested in the literature, as shown in section 8.2.3, that understanding an aural stimulus triggers participants to immediately initiate saccades to the target picture. By presenting such data I aim to show the extent of sensitivity to the inflection in the input.

Both reaction time and eye movement data come from the experimental trials that received a correct response only in the picture choice task. This is because in the trials that received incorrect response, it is assumed that the morphology was not perceived; hence, processing data in these trials cannot be considered a reflection of how participants process morphology.

Prior to presenting the results from these data, the experimental research hypotheses are presented below.

7.3 Research Hypotheses

As shown in the two previous chapters (Chapters five and six), the research question of this thesis is *What is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?* To answer this question, four hypotheses have been formulated to be tested against processing data in this chapter. These four hypotheses are parallel to the first four hypotheses in Chapters five (section 5.2) and Chapter six (section 6.2), which were set up based on the Organic Grammar (OG), Modulated Structure Building Hypothesis (MSBH), Representational Deficit Hypothesis (RDH) and Missing Surface Inflection Hypothesis (MSIH) accounts of morphological variability as well as the (dis)similarities between the L1s and L2 of the research participants and their L2 proficiency.

In formulating the four hypotheses for this study, it will be assumed that the same syntactic representations underlie production, perception and processing. Moreover, as elaborated in 7.2.3 above, following Jiang (2007), I will assume that automatic competence in L2 processing is parallel to mental representations in L2 acquisition; that is, developing an automatic competence means the same as building a mental representation. In Chapters five and six, consistency in production and accuracy in perception were taken as an indicator of the presence of underlying syntactic representations. In this chapter, it is quick processing that will be taken as indicator of automatic competence and thus presence of syntactic representations. Whether processing is quick or not will be judged against data collected from native speakers performing the same task. Bearing this in mind, the following four experimental hypotheses are set up based on the OG, MSBH, RDH and MSIH accounts, respectively.

H1. Arab and Chinese participants at all proficiency levels will process the morphology in a similar manner showing similar RTs, similar pattern of the length of eye looks at the target picture and similar speed in initiating the first look to the target picture. Both language groups will develop automatic competence with rising proficiency.

H2. Arab and Chinese learners at lower proficiency levels will process the morphology in a similar manner showing similar RTs, similar pattern of the length of eye looks at the target picture and similar speed in initiating the first look to the target picture. However, automatic competence will occur earlier for Arab than Chinese participants.

H3. The Arab participants at all proficiency levels will process the morphology more quickly than their Chinese counterparts, evidenced by shorter RTs, longer eye looks to target than competitor pictures and a quicker first eye-look to the target picture. Only Arab participants will exhibit automatic processing with Chinese participants lagging behind.

H4. The Arab and Chinese participants will similarly exhibit automatic competence at all proficiency levels, evidenced by quick RTs, longer eye looks to target than competitor pictures and a quick first eye-look to the target picture.

7.4 Results

In this section, results from RT data will be presented first in 7.4.1. This represents the speed at which participants respond to stimuli showing the automaticity of their access to knowledge. This is followed by the results from the two eye movement measures. The first eye-movement measure in 7.4.2 is the length of eye looks on a region of interest. This sheds light on to what extent learners prefer the target picture over the competitor one determined by the length of eye-looks on each. The second measure in 7.4.3 is the time of first look to target picture demonstrating the sensitivity to the morphology exhibited by the speed at which learners converge on the target picture. The gaze preference and gaze length data are claimed here to be aligned with data on reaction time and accuracy, but also to provide a more fine-grained insight into the cognitive processes of decision making than accuracy or RT on their own can provide.

As explained in the methodology section (7.2.5), the RT and eye-tracking data reported in this section are taken from the experimental trials that received a correct response only.

The data were coded into SPSS and tested for normal distribution using Shapiro-Wilk test. Many variables were found non-normally distributed and therefore non-parametric tests will be used for group comparisons.

The RT results are presented below.

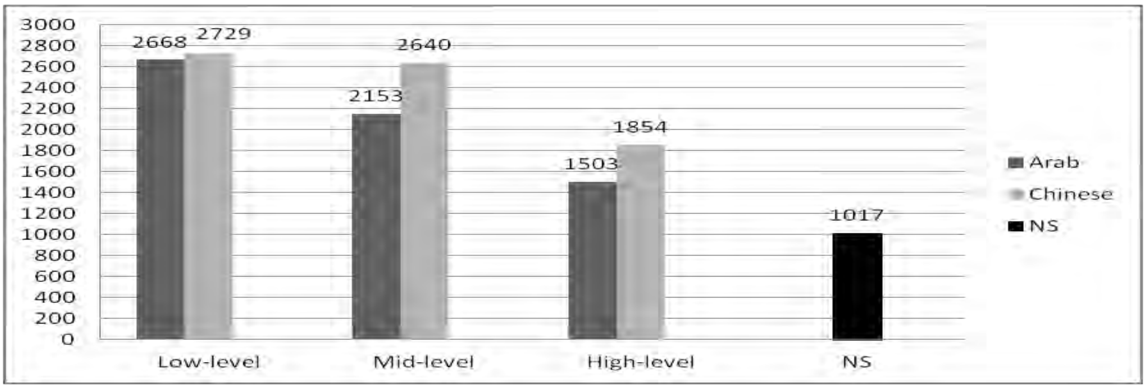
7.4.1 Reaction Times

Table 7.1 shows statistical data for RTs in response to past tense trials by non-native and native groups. Means are compared in Figure 7.1 to illustrate the differences between Chinese and Arab participants at every proficiency level and to compare them with the native speakers.

Table 7.1: RT (Millisecond) in past tense trials.

Groups	Mean	SD	Minimum	Maximum
Arabic Low level (n=9)	2668	1156	1046	4829
Arabic Mid level (n=13)	2153	878	887	4050
Arabic High-level (n=9)	1503	688	876	3024
Chinese Low level (n=11)	2729	687	1813	3787
Chinese Mid-level (n=13)	2640	1268	1499	5725
Chinese High-level (n= 10)	1854	1107	800	4656
Native speakers (n=10)	1017	210	774	1443

Figure 7.1: Mean RT scores (milliseconds) in past tense trials



Arab and Chinese participants' results show some differences at the three proficiency levels. First, to test for significance, Mann-Whitney tests were performed on the results of participants who are at the same proficiency level. The results showed the following

- In the past tense trials, there were no significant differences between the RTs of Arab and Chinese participants at any level (Low level: U=46, p=790); Mid level: U=62, p=.564; High level: U=39, p=.624)

Next, the development in RTs over proficiency was tested using Kruskal-Wallis tests performed on Arab and Chinese participants' scores separately. Results from the Arab groups showed the following:

- There was a significant difference among Arab different proficiency groups in their RTs to past tense trials ($\chi^2(2)=6.023$, $p<.05$, $\eta^2=.20$). Post hoc testing (Mann-Whitney) indicated that neither Low- and Mid-level groups nor Mid- and High-level groups differed from each other (Low vs. Mid: U=40, p=.320; Mid vs. High: U=29, p=.076, r=.38), but Low and High-level groups differed significantly (U=15, p<.05, r=.53).

Results from the Chinese groups showed the following:

- There was a significant difference among Chinese groups ($\chi^2(2)=6.309$, $p<.05$, $\eta^2=.21$). Post hoc testing (Mann-Whitney) indicated that the Low- and Mid-level learners performed similarly (U=49, p=.295) and so did Mid- and High-level learners (U=35, p=.099), but there was a significant difference between the Low-level and High-level learners (U=21, p<.05, r=.52).

-

In order to test how far non-native participants were similar to native speakers, Mann-Whitney tests comparing the control group with each of the non-native groups were performed. Results showed the following:

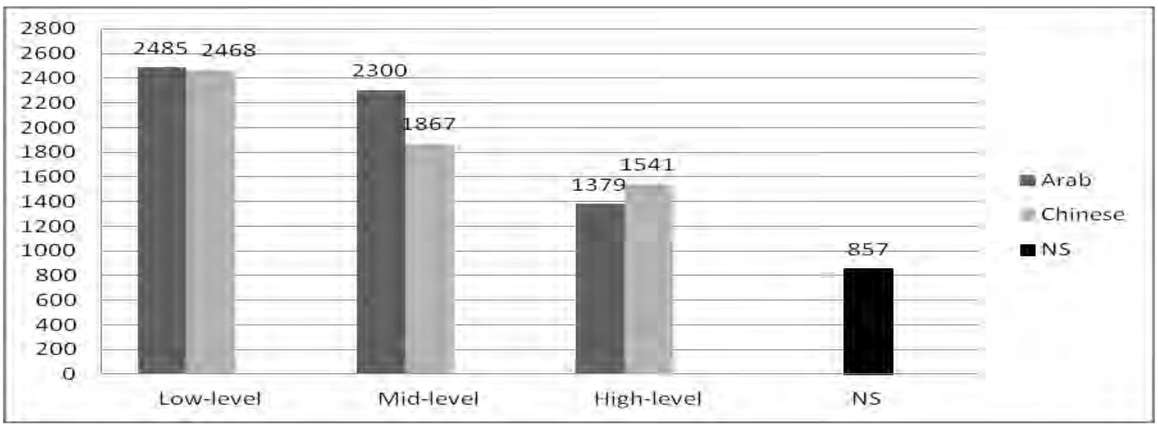
- Both Arab and Chinese groups at all proficiency levels performed significantly below the native speakers group in RTs to past tense trials (Arabic Low vs. NS: U=4, p<.005, r=.76; Arabic Mid vs. NS: U=10, p<.005, r=.70; Arabic High vs NS: U=17, p<.05, r=.52; Chinese Low vs. NS: U=.00, p<.001, r=.84; Chinese Mid vs. NS: U=.00, p<.001, r=.84; Chinese High vs. NS: U=22, p<.05, r=.47).

I turn now to RTs in response to the verbal agreement trials. Table 7.2 shows statistical data for RTs in response to verbal agreement trials by non-native and native groups. Means are compared in Figure 7.2 to illustrate the differences between Chinese and Arab participants at every proficiency level and to compare them with the native speakers.

Table 7.2: (Millisecond) in agreement trials.

Groups	Mean	SD	Minimum	Maximum
Arabic Low level (n=9)	2485	888	1709	4126
Arabic Mid level (n=13)	2300	1278	926	5482
Arabic High level (n=9)	1379	509	1000	2638
Chinese Low level (n=11)	2468	1069	1547	5411
Chinese Mid level (n=13)	1867	1015	693	4258
Chinese High level (n= 10)	1541	988	722	4142
Native speakers (n=10)	857	135	675	1085

Figure 7.2: Mean RT scores (milliseconds) in agreement trials



Arab and Chinese participants' results show some differences at the three proficiency levels. To test for significance, Mann-Whitney tests were performed on groups at the same proficiency level. The results showed the following

- In the verbal agreement trials, there were no significant differences between the RTs of Arab and Chinese learners at any level (Low level: U=48, p=.909; Mid level: U=66, p=.343; High level: U=42, p=.806).

The development in RTs with rising proficiency was tested using Kruskal-Wallis tests performed on Arab and Chinese participants' scores separately. Results from the Arab groups showed the following:

- There was a significant difference among different Arab proficiency groups in their RTs to verbal agreement trials ($\chi^2(2)=9.055$, $p<.05$, $\eta^2=.30$). Post hoc testing (Mann-Whitney) indicated that Low- and Mid-level groups were not statistically different from each other (U=50, p=.570) but both performed significantly below the High-level group (High vs. Low: U=6, p<.01, r=.71; High vs. Mid: U=26, p<.05, r=.46).
-

Results from the Chinese groups showed the following:

- There was a significant difference among Chinese groups ($\chi^2(2)=7.585$, $p<.05$, $\eta^2=.22$). Post hoc testing (Mann-Whitney) indicated that that neither the Low- and Mid-level groups nor the Mid- and High-level groups were statistically different from each other (Low vs. Mid: U=43, p=.099, r=.33; Mid and High: U=49, p=.321) but that the Low-level group performed significantly below the High-level group (U=15, p<.01, r=.61).
-

In order to test the difference between non-native and native groups, Mann-Whitney tests were performed. Results showed the following:

- Both Arab and Chinese groups at all proficiency levels performed significantly below the native speakers group in RTs to verbal agreement trials (Arabic Low vs. NS: U=.00, p<.001, r=.84; Arabic Mid vs. NS: U=2, p<.001, r=.81; Arabic High vs. NS: U=3, p<.005, r=.78; Chinese Low vs. NS: U=.00, p<.001, r=.84;

Chinese Mid vs. NS: *U*=12, p<005, r=.68; Chinese High vs. NS: *U*=19, p<.05, r=.52).

To sum up, the RT data showed that there were no differences between Arab and Chinese learners at any of the Low, Mid or High proficiency levels. It also demonstrated a remarkable improvement with rising proficiency, with significant differences for both Arab and Chinese participants alike found between Low/ Mid and High proficiency groups. However, Arab and Chinese groups at the three proficiency levels performed significantly below the control group.

Therefore, similar to the perception results, the RT measure did not show statistical differences between the two L1 groups. Recall that the statistical tests on perception, comparing Arab and Chinese groups at the three levels showed no statistical difference in any cases apart from the High-level group where the Arab group performed better than its Chinese counterpart in the perception of verbal agreement at an approaching-significance level. However, different from perception results, which showed a pattern of improvement with rising proficiency among the Arab participants and no improvement among the Chinese participants, the RT results demonstrated a developmental trajectory of remarkable improvement (shorter RTs) with rising proficiency for Arab and Chinese participants alike. Also in perception of both past tense and verbal agreement, while Arab participants at the High level performed in a native-like manner, Chinese participants at the same level performed significantly below native speakers. In processing, however, both Arab and Chinese High level participants performed significantly below native speakers.

The eye-tracking data are presented hereafter.

7.4.2 Length of Eye Looks: Target vs. Competitor

I start at first by presenting the length of eye looks at target vs. competitor pictures in past tense trials and then in agreement trials for native and non-native groups. Figure 7.3 illustrates the eye looks of native speakers group.

Figure 7.3: Eye-looks in past tense trials by NS

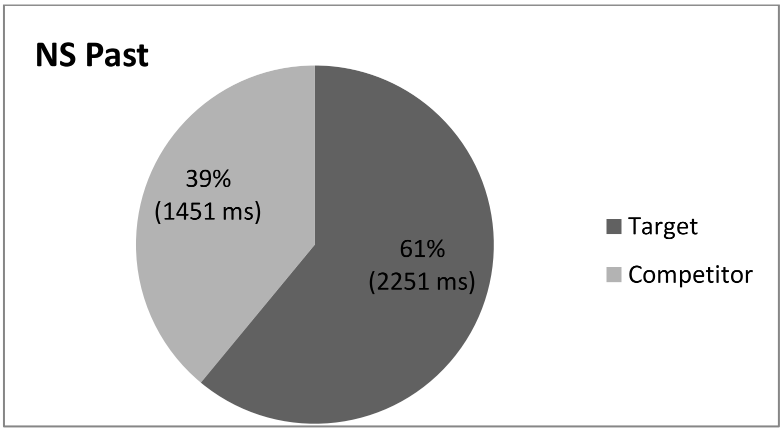


Figure 7.3 shows that the control group looked at the target picture 61% of the total time and at the competitor picture 39% of the total time. A Wilcoxon signed rank test comparing target and competitor looks showed that:

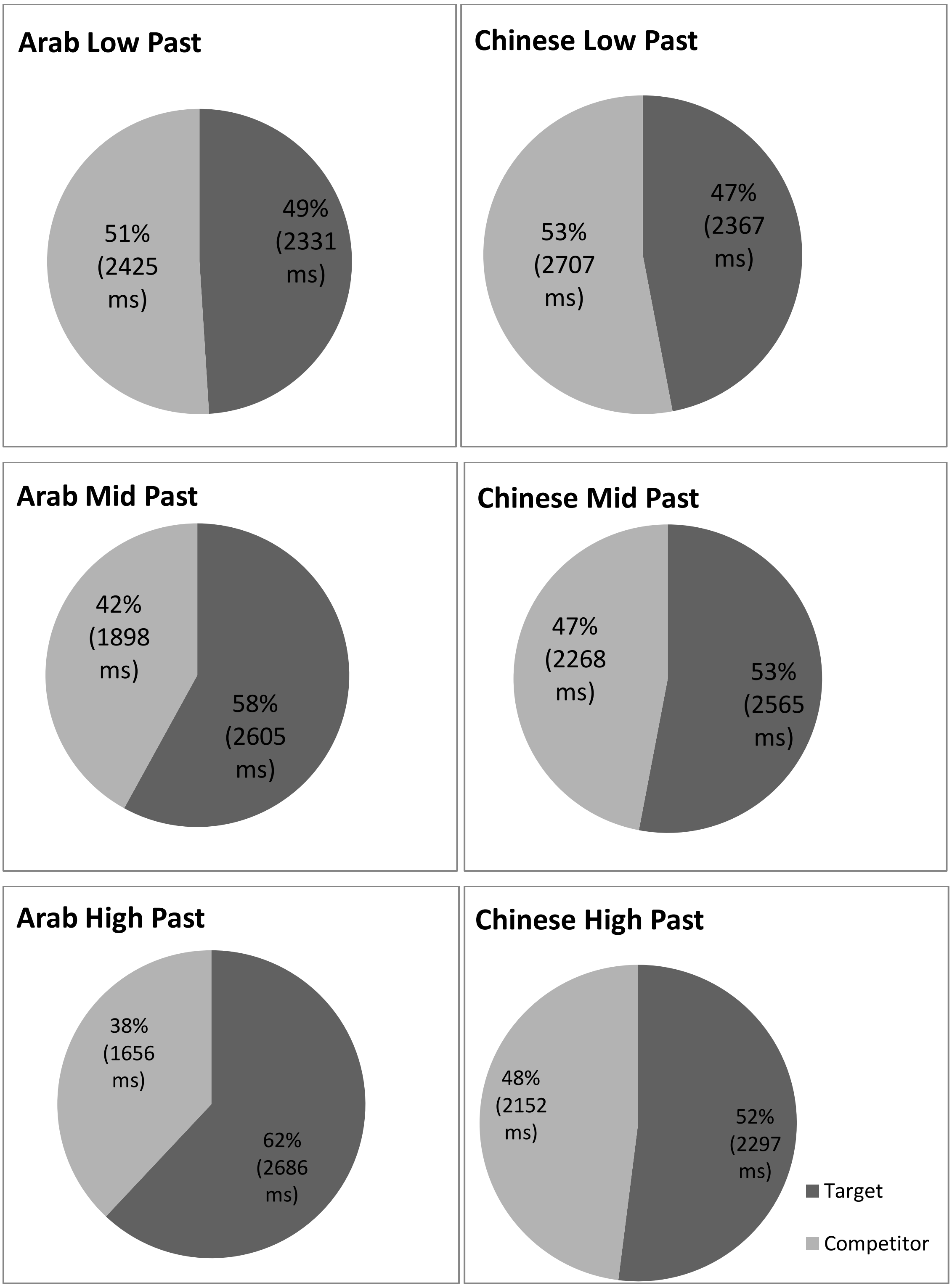
- The native speakers group looked significantly at the target picture more than the competitor picture in the past tense trials (Z= -2.701 p=.01, r=.60).

Table 7.3 summarises the length and proportion of time spent on a target and competitor pictures by all non-native groups in past trials and Figure 7.4 gives pie chart illustrations.

Table 7.3: Eye-looks (mean lengths in proportions) in past tense trials by non-native groups

	Target Looks		Competitor Looks	
	Mean Length	%	Mean Length	%
Arab Low	2331 ms	49	2425 ms	51
Arab Mid	2605 ms	58	1898 ms	42
Arab High	2686 ms	62	1656 ms	38
Chinese Low	2376 ms	47	2707 ms	53
Chinese Mid	2565 ms	53	2268 ms	47
Chinese High	2297 ms	52	2152 ms	48

Figure 7.4: Eye-looks in past tense trials by non-native groups

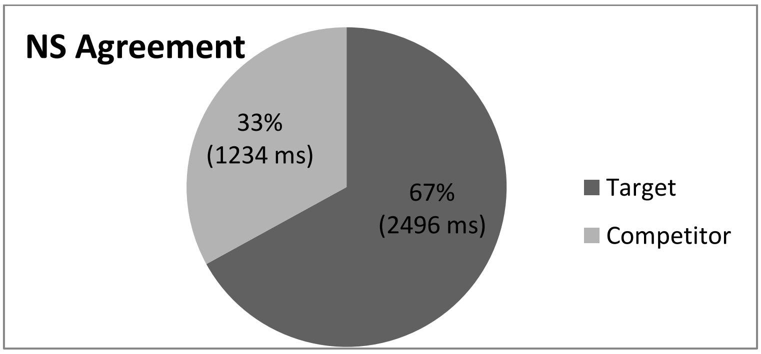


To check which groups looked reliably more to target pictures, Wilcoxon signed rank tests comparing target and competitor looks were performed on the mean score of each group separately. These showed the following:

- At the Low level, Arab and Chinese participants looked at the target and competitor pictures in the same manner showing no significant preference for either of them ((Z=-.059, p=.953) for Arab participants and (Z = -.889, p=.386) for Chinese participants).
- At the Mid level, the Arab participants looked significantly at the target picture more than the competitor picture (Z=-2.510, p<.05, r=.51). By contrast, the Chinese participants did not show preference to either picture (Z=-1.255, p=.209)..
- At the High level, while the Arab participants looked significantly at the target picture more than the competitor picture (Z=-2.547, p<.05, r=.60), their Chinese counterparts showed no preference for either of them (Z=-.866, p=.386).

We now turn to the performance of the same participants on verbal agreement trials. We start first with the native groups' results in Figure 7.5.

Figure 7.5: Eye-looks in agreement trials by NS



The pie chart shows that the control group looked at the target picture 67% of the total time and at the competitor picture 33% of the total time. A Wilcoxon signed rank test comparing target and competitor looks showed the following:

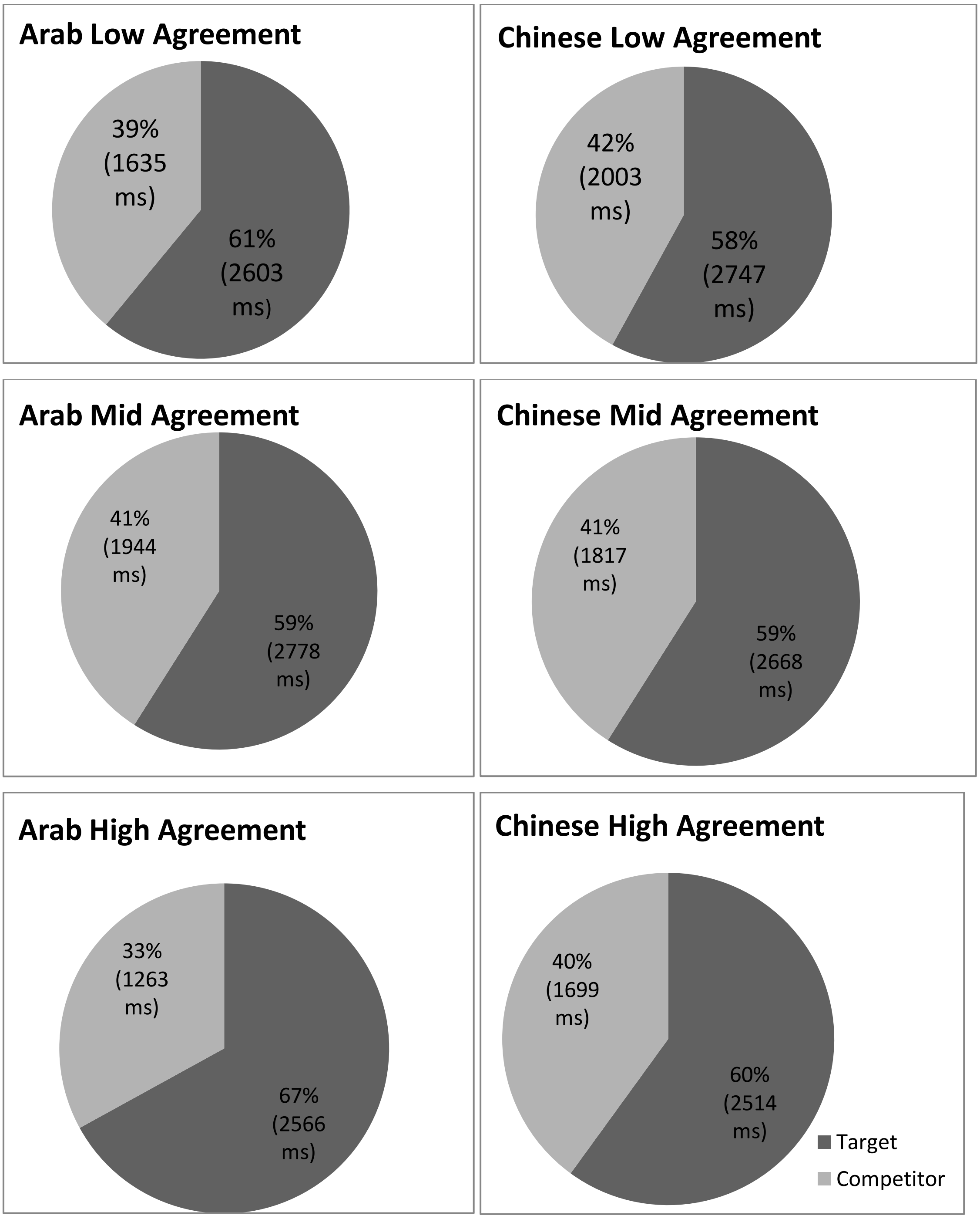
- The native speakers group looked significantly at the target picture more than the competitor picture in verbal agreement trials (Z=-2.803, p<.01, r=.62)

We move next to the non-native group results. The length and proportion of time spent on a target and competitor pictures in verbal agreement trials are summarised in Table 7.4 and pie chart illustrations are given in Figure 7.6.

Table 7.4: Eye-looks (mean lengths in proportions) in verbal agreement trials by non-native groups

	Target Looks		Competitor Looks	
	Mean Length	%	Mean Length	%
Arab Low	2603 ms	61	1635 ms	39
Arab Mid	2778 ms	59	1944 ms	41
Arab High	2566 ms	67	1263 ms	33
Chinese Low	2747 ms	58	2003 ms	42
Chinese Mid	2668 ms	59	1817 ms	41
Chinese High	2514 ms	60	1699 ms	40

Figure 7.6: Eye-looks in agreement trials by non-native groups



Wilcoxon signed rank tests comparing target with competitor looks were performed on the results of each non-native group separately. These showed the following:

- At all proficiency levels, Arab and Chinese participants looked significantly at the target picture more than the competitor picture in verbal agreement trials (Arab Low: Z= -2.310, p<.05, r=.54; Chinese Low: Z= -2.312, p<.05, r=.49; Arab Mid: Z=-2.830, p<.01, r=.55; Chinese Mid: Z=-2.746, p<.01, r=.53; Arabic High: Z=-2.666, p<.01, r=.62; Chinese High: Z=-2.090, p<.05, r=.46)

All in all, the length of eye looks measure showed that in agreement trials both Arab and Chinese participants at all proficiency levels looked more reliably to the target pictures than the competitor ones. In past tense trials, however, the Arab participants at the Mid and High levels looked reliably more to the target than competitor ones whereas the Chinese Low, Mid and High-level learners showed no preference for either of the two pictures.

The sensitivity measure results are presented next.

7.4.3 Sensitivity to Morphology: The First Look to Target

This section presents the results of another eye-tracking measure providing information on how the research participants processed the morphology under investigation. In this measure, the time of the first look to the target picture after the presentation of the verb inflected for past tense or verbal agreement is considered an indicator of the participants' sensitivity to the information inherent in the morphological items.

Table 7.5 shows statistical data for the time of the first look at the target picture in past tense trials by non-native and native groups. Means are compared in Figure 7.7 to illustrate the development over proficiency, the differences between Chinese and Arab participants at every proficiency level and between all non-native groups and the native speakers.

Table 7.5: Sensitivity (Millisecond) in past tense trials.

Groups	Mean	SD	Minimum	Maximum
Arabic Low level (n=9)	1246	426	828	2021
Arabic Mid level (n=13)	940	270	339	1410
Arabic High level (n=9)	773	395	467	1695
Chinese Low level (n=11)	1654	367	994	2217
Chinese Mid level (n=13)	1126	459	413	1810
Chinese High level (n= 10)	1425	528	494	1969
Native speakers (n=10)	695	243	284	1081

Figure 7.7: Means of the time of the first look (in millisecond) at the target picture in past tense trial by native and non-native groups.

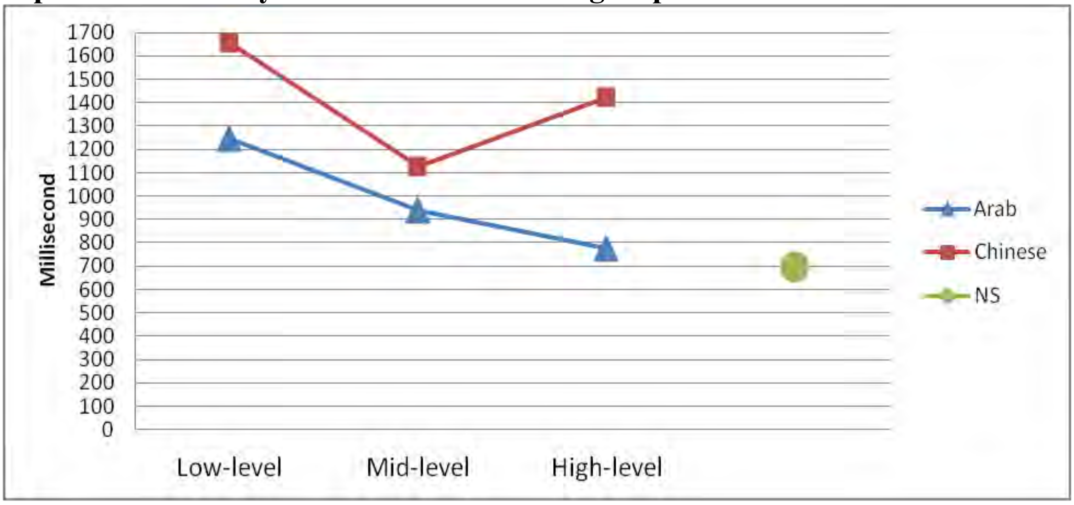


Figure 7.7 shows that the Arab groups' sensitivity to past morphology appears to have increased with higher proficiency. It also shows that the Chinese groups' sensitivity to past tense morphology appears to have increased at the Mid level but then decreased at the High level. To determine the significance of the development from one group to a higher proficiency group, a Kruskal-Wallis test was performed. The results of the Chinese groups showed the following:

- There was a significance difference among Arab different proficiency groups in their sensitivity to past tense morphology ($\chi^2(2)=8.295$, $p<.05$, $\eta^2=.27$). Post hoc test (Mann-Whitney) indicated that there were no significant differences between the Low and Mid-level groups ($U=34$, $p=.102$), but the High-level group's sensitivity was higher than that of the Low-level and Mid-level groups, significantly for the former ($U=11$, $p<.01$, $r=.61$) and at the approaching significance level for the latter ($U=31$, $p=.066$, $r=.39$).

The results of the Chinese groups showed the following:

- There was an approaching-significance difference among Chinese different proficiency groups in their sensitivity to past tense morphology ($\chi^2(2)=5.727$, $p=.057$, $\eta^2=.17$). Post hoc test (Mann-Whitney) indicated that the Mid-level group's sensitivity was higher than that of the Low-level group ($U=29$, $p<.05$, $r=.47$), but there were no significant differences between either the Low and High-level groups ($U=40$, $p=.470$) or the Mid and High-level groups ($U=32$, $p=.118$).

To check how different-L1 non-native speakers at the same proficiency level differ from each other and from native speakers, Mann-Whitney tests were performed. The results show the following:

- There were significant differences between the Arab and Chinese groups at the Low level ($U=23$, $p<.05$, $r=.45$) and High-level ($U=13$, $p<.05$, $r=.57$), with the Arab participants showing higher sensitivity. At the Mid-level, however, there was no significant difference between Arab and Chinese participants ($U=61$, $p=.376$).
- There were significant differences between native speakers and Arab groups at Low ($U=10$, $p<.01$, $r=.65$) and Mid levels ($U=32$, $p<.05$, $r=.42$), but no significant difference between the control group and the Arab High-level group was detected ($U=43$, $p=.902$).
- There were significant differences between the control group and the Chinese groups at all proficiency levels (Chinese Low vs. NS: $U=1$, $p<.001$, $r=.82$; Chinese Mid vs. NS: $U=28$, $p<.05$, $r=.44$; Chinese High vs. NS: $U=11$, $p<.005$, $r=.63$).

I turn now to the results of the agreement trials.

Table 7.6 shows statistical data for the time of the first look at the target picture in verbal agreement trials by non-native and native groups. Means are compared in Figure 7.8 to illustrate the development with rising proficiency, the differences between Chinese and Arab participants at every proficiency level and between all non-native groups and the native speakers.

Table 7.6: Sensitivity (Millisecond) in verbal agreement trials.

Groups	Mean	SD	Minimum	Maximum
Arabic Low level (n=9)	1109	242	800	1592
Arabic Mid level (n=13)	933	496	379	2356
Arabic High level (n=9)	741	197	377	1104
Chinese Low level (n=11)	1195	178	1007	1532
Chinese Mid level (n=13)	878	350	366	1443
Chinese High level (n= 10)	1046	342	708	1798
Native speakers (n=10)	589	190	299	884

Figure 7.8: Means of the time of the first look (in millisecond) at the target picture in verbal agreement trial by native and non-native groups.

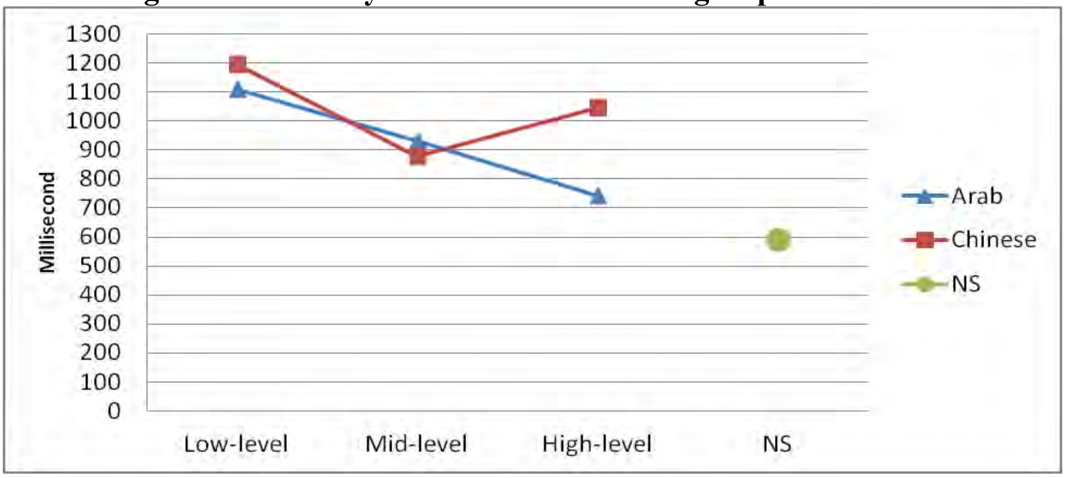


Figure 7.8 shows that the Arab groups' sensitivity to the verbal agreement inflection appears to have increased with rising proficiency. It also shows that the Chinese groups' sensitivity to appears to have increased at the Mid level but then decreased at the High level. To determine the significance of the development of sensitivity with rising proficiency, a Kruskal-Wallis test was performed. The results of the Arab groups showed the following:

- There was a significant difference among Arab different proficiency groups in their sensitivity to verbal agreement inflection ($\chi^2(2)=8.358$, $p<.05$, $\eta^2=.27$). Post hoc testing (Mann-Whitney) indicated that the Mid-level and High-level groups performed significantly better than the Low-level group (Mid-level vs. Low-level: $U=29$, $p<.05$, $r=.42$; High-level vs. Low-level: $U=7$, $p<.005$, $r=.69$) but the difference between the Mid-level and High-level groups was not significant ($U=47$, $p=.443$).

The results of the Chinese groups showed the following:

- There was a significant difference among Chinese different proficiency groups in their sensitivity to verbal agreement inflection ($\chi^2(2)=5.976$, $p<.05$, $\eta^2=.18$). Post hoc testing (Mann-Whitney) indicated that the Mid-level group performed significantly better than the Low-level group ($U=33$, $p<.05$, $r=.45$), but the difference between the Mid-level and High-level groups was not significant ($U=41$, $p=.243$) and so was the difference between the Low-level and High-level groups ($U=29$, $p=.119$).

These results demonstrate that Arab and Chinese groups showed similar patterns of development as there were significant differences between the Low-level and Mid-level groups and no statistically significant differences between the Mid-level and High-level groups. However, although there were no statistical differences between Mid-level and High-level groups for Arab and Chinese participants alike, while the Arabic speakers' sensitivity to morphology increased at the High-level as they showed descriptively better performance than the Mid-level participants, the Chinese speakers' sensitivity decreased at the High-level as appears when compared to the Mid-level participants.

To check how different-L1 non-native speakers at the same proficiency level differ from each other and from native speakers, Mann-Whitney tests were performed. The results show the following:

- There was no significant difference between the Arab and Chinese groups at either the Low level ($U=37$, $p=.342$) or Mid level ($U=79$, $p=.778$), but the Arab High-level participants' sensitivity to the agreement inflection was significantly higher than the sensitivity of their Chinese counterparts ($U=17$, $p<.05$, $r=.48$)
- There were significant differences between native speakers and Arab groups at Low and Mid levels (Arab Low vs. NS: $U=2$, $p<.001$, $r=.80$; Arab Mid vs. NS: $U=25$, $p<.05$, $r=.51$), but no significant difference between the control group and the Arab High-level group was detected ($U=24$, $p=.094$, $r=.38$).
- There were significant differences between the control group and the Chinese groups at all proficiency levels (Chinese Low vs. NS: $U=.00$, $p<.001$, $r=.84$; Chinese Mid vs. NS: $U=33$, $p<.05$, $r=.41$; Chinese High vs. NS: $U=8$, $p<.005$, $r=.69$).

All in all, in both past tense and verbal agreement trials, the speed of the first look to the target picture showed some differences between Arab and Chinese participants. While Arab participants' sensitivity increased with rising proficiency at either the Mid or High levels, the Chinese participants' sensitivity increased at the Mid level and then decreased at the High level. Moreover, Arab participants' speed of initiating the first look to the target picture at the High proficiency level was significantly higher than the speed of Chinese participants. Comparisons between non-native and native groups revealed that only the Arab High level group performed similar to native speakers and all other groups lagged behind.

7.5 Summary and Hypothesis Evaluation

In this section the results of the experiment are summarised first. Then, prior to proceeding to hypotheses evaluation, whether non-native participants developed automatic competence will be judged based on their performance relative to that of native speakers.

First, the RTs of Arab and Chinese participants were similar at all proficiency levels in both past tense and verbal agreement trials. The two language groups showed shorter RTs with rising proficiency but all non-native groups performed significantly below native speakers.

Second, the gaze length measure showed that in verbal agreement trials there were no differences in the eye-look patterns between Arab and Chinese participants at any proficiency level; however, in past tense trials, differences emerged as only Arab participants at the Mid and High levels looked more reliably to the target picture, with the Chinese participants at all proficiency levels showing no preference to any of the target or competitor pictures.

Third, the sensitivity measure showed that the Arab Low and Mid-level participants' speed in looking to the target picture in both verbal agreement and past tense trials was similar to that of their Chinese counterparts, but at the High level, the Arab group performed significantly faster than the Chinese and only the Arab were similar to native speakers with the Chinese performing significantly below.

Concerning automatic competence, as noted in 7.2.3, this will be judged based on the non-native speakers' performance relative to that of native speakers. Chinese participants at the High level were similar to native speakers only in eye gaze patterns to pictures in verbal agreement trials, differing from them in eye gaze patterns in past trials, RTs in past and agreement trials and speed of first look to target in past and agreement trials. Arab speakers at the High level differed from native speakers in RTs to past and agreement trials but they were similar in eye gaze patterns and speed of eye look to target picture in past and agreement trials. Given that the RT is the time between the end of the full verbal stimulus and the participants' decision and, by contrast, the speed of first eye look is the time between the end of the inflected verb and the first look to target picture, it can be assumed that the latter type of data provides a more direct window into investigating the automaticity of the processing of the properties under study. This will be discussed further in the Chapter eight, but based on this for the purpose of evaluating the experimental hypotheses, I maintain that Arab, but not Chinese, High level participants developed automatic competence.

As for the research hypotheses for the present experiment, they will be repeated below and each will be followed with an evaluation of its power in predicting the attested performance.

> *H1. Arab and Chinese participants at all proficiency levels will process the morphology in a similar manner showing similar RTs, similar pattern of the length of eye looks at the target picture and similar speed in initiating the first look to the target picture. Both language groups will develop similarly automatic competence with rising proficiency.*

This hypothesis (based on OG) is partially confirmed. What is consistent with this hypothesis is that Arab and Chinese participants at Low and Mid levels performed similarly in RTs, length of eye gazes and first look to target with the exception of the length of eye gazes in past items, where only Arab speakers looked longer at target pictures. However, the performance of the High-level groups constitutes a challenge for this hypothesis. First, Arab, but not Chinese, speakers looked longer to target picture in past tense trials and their first looks to target picture in both past and agreement trials were significantly faster. Second, only Arab speakers developed automatic competence with the Chinese participants falling behind.

H2. Arab and Chinese learners at lower proficiency levels will process the morphology in a similar manner showing similar RTs, similar pattern of the length of eye looks at the target picture and similar speed in initiating the first look to the target picture. However, automatic competence will occur earlier for Arab than Chinese participants.

This hypothesis (based on the MSBH) is widely supported. Arab and Chinese participants' participants patterned similarly at Low and Mid levels. Also, the data showed that only Arab participants developed automatic competence.

H3. The Arab participants at all proficiency levels will process the morphology more quickly than their Chinese counterparts, evidenced by shorter RTs, longer eye looks to target than competitor pictures and a quicker first eye-look to the target picture. Only Arab participants will exhibit automatic processing with Chinese participants lagging behind.

This hypothesis (based on the RDH) is partially supported. The performance of the High level participants which showed an advantage for Arabs over the Chinese gives support to this hypothesis. However, that Arabs and Chinese speakers patterned similarly at Low and Mid levels contradicts the prediction of this hypothesis.

H4. The Arab and Chinese participants will similarly exhibit automatic competence at all proficiency levels, evidenced by quick RTs, longer eye looks to target than competitor pictures and a quick first eye-look to the target picture.

This Hypothesis (based on the MSIH) is not supported. All non-native groups, apart from the Arab High level group, performed significantly below native speakers mainly in RTs and first look to target pictures.

The following chapter will discuss the findings reported in this and the previous two chapters (five and six) in the light of previous research reviewed in Chapters two and three.

Chapter 8. Discussion and Conclusions

8.1 Introduction

This chapter discusses the results from the production, perception and processing studies (Chapters five, six and seven) in the light of the literature reviewed in Chapters two and three and in connection with the main research question presented in Chapter one and repeated hereafter.

What is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?

I will start with the summary and discussion of findings in section 8.2. Providing an answer to the main research question is attempted in section 8.3. Limitations of the studies and recommendations for further research are discussed in 8.4 and the chapter is concluded in 8.5.

8.2 Summary and Discussion of Findings

The main goal of this thesis is to locate the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in the speech of L2ers of English. To this end, I conducted three studies on the production, perception and processing of these properties by 71 native speakers of Arabic (n=34) or Mandarin Chinese (n=37). In order to make comparisons between the performance of the participants from the two language backgrounds, they were placed into comparable groups according to their L2 proficiency, which was measured by ASCOPS (Unsworth 2005, 2008).

ASCOPS (chapter four) was performed on semi-spontaneous production data collected from the research participants using a picture description task. This placed them at three proficiency levels: Low, Mid and High. One very important advantage of using ASCOPS was that it made it possible to assess L2 proficiency regardless of the status of the relevant morphological paradigm in the L2ers' speech. I argued that past tense and verbal agreement morphology should not be a determining factor in assessing the proficiency of the participants in this research. The rationale for this was to avoid ending up placing the participants in proficiency groups relative to each other according to their performance on the same properties that the main experiments tested. Therefore, by using ASCOPS, a confounding factor was avoided.

In the production study (Chapter five), data were collected using a sentence elicited imitation task. Although this method has been criticised as being subject to measuring participants' ability to imitate the stimulus rather than measuring their linguistic competence, I argued that a good task design could obviate this problem. Following recent developments in the Elicited Imitation (EI) technique, I introduced a new procedure to the task used in this study. The new procedure was presenting a set of three pictures on a computer screen along with the oral stimulus. Participants had to choose one picture before repeating the oral stimulus. Such a procedure forced the participants to parse the stimulus for comprehension and thus prevented a verbatim imitation.

In the perception and processing studies (Chapters six and seven, respectively), a computerised picture-choice task supplemented with RT and eye-movement measures was designed. The task presented three pictures along with an aural sentence and participants were asked to choose the picture that best depicted the oral stimulus. The choice of the picture depended on the participants' perception of the past tense and verbal agreement morphology. On the one hand, the picture choice response provided accuracy results on whether the morphology was perceived or not. The RT and eye-movement data, on the other hand, revealed online information on how the morphological items were processed and to what extent participants were sensitive to their presence in the stimuli.

Furthermore, as detailed in chapters five and six, in the two main tasks (the sentence elicited imitation and computerised picture choice tasks), the phonological form of inflected verbs for past or agreement was controlled for; that is, balanced number of verbs which end in consonant clusters and non–clusters was included. Irregular verbs in past tense items were also tested. This was in the aim of examining whether the occurrence of the inflection in phonologically complex codas is the source of the production or perception problems.

In the following three sub-sections, the results of the production, perception and processing studies will be summarised and discussed. Prior to this, it is worth reminding the reader with the linguistic assumptions on the participants' native languages regarding the properties under study. The native languages of the participants (Arabic and Mandarin Chinese) resemble and differ from each other and from the target language (English) in important ways concerning the relevant morphosyntactic and morphophonological characteristics. These are summarised as follows. It has been assumed that Arabic is similar to English in having syntactic features for past and verbal agreement, but Chinese differs from both as it lacks them. With regard to the consonant cluster structure created by the addition of the English *–ed* and *–s* inflections to verb stems of specific shapes, it has been demonstrated that only Arabic allows it in coda position, with Mandarin Chinese disallowing it in any word position. It has been also assumed that the English inflection is prosodified PWd externally violating EXHAUST and NONREC simultaneously. This is in contrast to both Arabic and Chinese, neither of which contains structures prosodified in such a fashion.

8.2.1 Results of the Production Study

The results of the production study (Chapter five) showed that morphological variability in the use of past and agreement was experienced by Chinese participants at Low, Mid and High proficiency levels and by Arab participants at the two lower levels. This phenomenon was not attested in the performance of Arab High level informants. Statistical tests revealed that the two language groups did not differ from each other at the Low or Mid levels, but a significant difference was detected at the High level with the advantage to the Arab participants. It was also found that the high level participants in the two language groups supplied the morphology significantly better than their lower proficiency peers.

Pursuing the source of variability at the two lower proficiency levels and the advantage that Arabic speakers at the High proficiency level had over their Chinese counterparts, the results were broken down by allomorph type (C-d/z, C-t/s, VV-d/z and CV-d/z), cluster type (consonant-cluster and non-cluster), verb type (regular and irregular) and prosodic structure (PWd internal and PWd external). For convenience, the findings of these analyses are summarised in Tables 8.1 for past tense and 8.2 for verbal agreement.

Table 8.1: A summary of the results of phonological analyses on past tense morphology production by all non-native groups

	Allomorphic variants	Consonant cluster vs non-cluster	Prosodic Word: Internal vs external	Regular vs irregular
Arab Low	All allomorphs variable C-t Lower than C-d	No difference	PWd internal: all structures variable Internal vs external: comparison failed*	Both variable No difference
Chinese Low	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: comparison failed*	Both variable Irreg higher than Reg
Arab Mid	All allomorphs variable C-t Lower than C-d and VV-d	Consonant cluster Lower than Non-consonant cluster	PWd internal: all structures variable Internal vs external: comparison failed*	Both variable No difference
Chinese Mid	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: comparison failed*	Both variable Irreg higher than Reg
Arab High	All allomorphs accurate No differences	No difference	PWd internal: all structures accurate Internal vs external: comparison failed*	Both accurate No difference
Chinese High	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: comparison failed*	Both variable No difference

* The number of contexts where the inflection was assumed to prosodify word externally was small, which prevented comparison with other structures.

Table 8.2: A summary of the results of phonological analyses on verbal agreement inflection production by all non-native groups

	Allomorphic variants	consonant Cluster vs non-cluster	Prosodic Word: Internal vs external
Arab Low	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: No difference
Chinese Low	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: No difference
Arab Mid	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: No difference
Chinese Mid	All allomorphs variable No differences	No difference	PWd internal: all structures variable Internal vs external: No difference
Arab High	All allomorphs accurate No differences	No difference	PWd internal: all structures accurate Internal vs external: No difference
Chinese High	All allomorphs variable CV-z higher than VV-z	No difference	PWd internal: all structures variable Internal vs external: No difference

As for the research question addressed in this thesis on the source of variability and the reason for its persistence in some learners rather than others, what the production study findings can tell us is that (L1-transferred) phonological constraints, as proposed by Lardiere (1998a, 2003), Goad et al. (2003) and Goad and White (2006), are not the answer. The phonological analyses summarised in Tables 8.1 and 8.2 showed some effects that might be argued to be phonological. Yet, these centred at the Low and Mid levels where no difference in the overall scores between Arab and Chinese groups was found. These effects are rather marginal in the sense that they cannot explain variability in Low and Mid levels or the difference between language groups at the High level. These results came from participants who speak two different languages one of which is similar to English as it allows consonant clusters (Arabic) and the other is different from English as it disallows consonant clusters (Mandarin). Despite this difference, the participants at Low and Mid levels did not differ in their overall L2 performance, showing similarly a severe variability. This clearly indicates that the variability experienced by these participants cannot be associated with L1-transferred phonological constraints of this kind. Moreover, the difference between the Arab and Chinese High level groups cannot either be explained by L1 transferred phonological constraints. This is because no phonological effects were observed at this level.

The most prominent effect observable in Table 8.1 is the higher production of morphology in irregular than regular verbs in past tense contexts by Chinese Low and Mid level participants. This might be interpreted as that there are phonological constraints on the production of the inflection. A closer look at the results, however, renders this interpretation unsupported. Although past production is higher in irregular than regular verbs, the Chinese informants' production rates were 55% for Low-level participants and 61% for Mid-level participants, showing robust variability. If the difference were due to phonological constraints, higher rates of past production in irregular verbs should have been observed. A probably more viable interpretation of the difference in past production between regular and irregular verbs is associated with the nature of these participants as being instructed learners. In classrooms, irregular verbs are taught as exceptions to the rule 'add *–ed* to the verb in past tense context' and learners recognize from early on that these exceptions should be memorised. Memorising the past tense forms of irregular verbs might lead to their being stored as separate entities in the lexicon of learners. Accordingly, this might make the task of recalling these forms easier than applying the regular verb rule, especially in a sentence repetition task in which the past tense form is provided.

Another observed phonological effect (but which does not explain variability) was the production of past *–ed* in non-cluster more than cluster contexts by Arab Mid level participants. Looking at the results broken down by allomorph, it can be seen that this was because the C-t (e.g., asked /kt/) allomorph was produced significantly lower than the VV-d (e.g., played /eid/) allomorph. The C-t allomorph was also produced lower than the C-d (e.g., learned /nd/) allomorph by Arabic speakers at the Low and Mid levels. Therefore, in addition to Arab participants' general problem with past tense inflection, it is clear that they had a difficulty with the production of the C-t allomorph. This difficulty, however, was overcome at the highest proficiency level.

That no phonological effects on the results of Mandarin participants was found might seem surprising at first glance. However, this is not inconsistent with what has been previously reported in the literature. The evidence found in previous studies for the involvement of L1-transferred phonological constraints is not compelling. One such evidence came from Lardiere (2003; see section 3.2.1). The evidence advanced by Lardiere on constraints at the syllable level was that 1) Patty's emails had higher rates of inflection suppliance than what was found in her speech and 2) Patty's production of past tense morphology was higher on irregular than regular verbs. This was taken by Lardiere as evidence for the presence of phonological constraints on the production of the inflection, especially in word-final consonant clusters. However, as argued in Chapter 3, Patty's better performance in her emails might rather be driven by a reliance on metalinguistic knowledge, which could more likely be employed during writing emails than free speech. Also, the difference between the production of past on regular and irregular verbs is not convincing evidence for the role of phonological constraints. In the production study in this thesis, a similar difference between the two types of verbs was also observed, yet the comparisons between the production of the inflection in consonant cluster and non-cluster contexts revealed no difference. This indicated that constraints on consonant clusters were not involved. Therefore, the results in this production study are not inconsistent with Lardiere's study results, but it is the results of the more detailed analysis conducted here that posed a challenge for Lardiere's interpretation

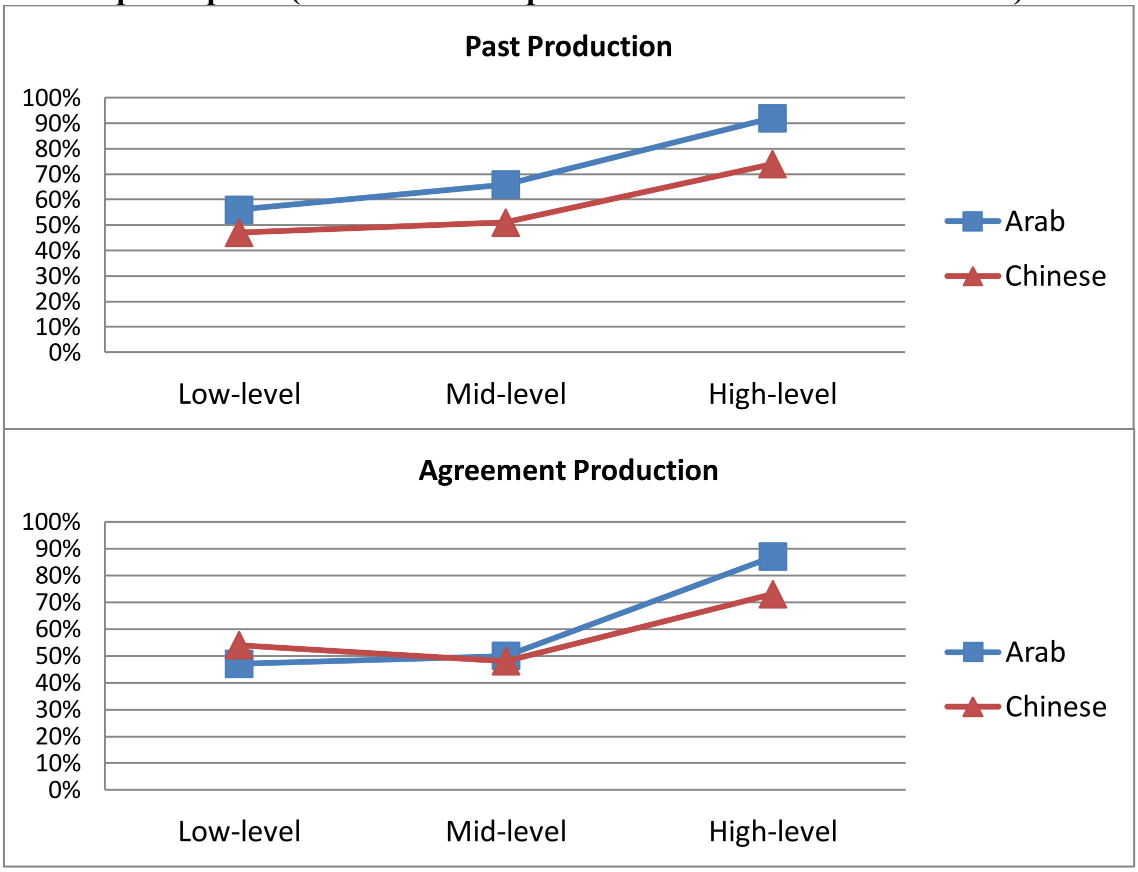
Another evidence for the involvement of phonological constraints is found in Goad et al. (2003; see section 3.2.1). This evidence was not conclusive either. Goad et al.'s data came from only six Mandarin-speaking L2 learners of English. The authors found that some prosodic structures were more difficult than others for their high-intermediate/ low-advanced informants. However, in another study by the first two authors (i.e., Goad and White, 2006), other 12 learners from the same L1 were not found to have the same difficulty although they were lower in L2 proficiency than Goad et al.'s informants. This might indicate that the results from the six informants in Goad et al.'s study are not generalisable. Indeed, this was not attested in this thesis production study as the results from the comparisons between different prosodic structures revealed that variability affected all structures similarly.

Then, what causes variability? The findings of the present production study indicate that the MSIH (Prévost and White, 2000b) cannot explain the observed performance. The MSIH is a Full Transfer/ Full Access account. This means that syntactic representations will be available to L2ers from their L1 or UG. Relating this to the present experiment, it implies that both Arab and Chinese participants will instantiate the relevant representations for past and agreement from early on (possibly from different sources though). Moreover, the performance of the two language groups clearly shows that they have knowledge of the surface form from the low level onward (probably by virtue of being instructed learners). Hence, according to the MSIH, it is a processing failure in matching the underlying syntax with the surface form that causes the variable production. However, the Arab and Chinese participants at Low and Mid levels did not show differences, but at the High level, thy diverged significantly. This is inconsistent with the MSIH because it is not clear how a mere processing failure would affect two language groups similarly at one level and then generate differences at a higher level.

Then, is variability due to a deficit in the underlying representations in line with RDH (Hawkins and Liszka, 2003)? The performance of the High level participants in this production study is consistent with what is reported by Hawkins and Liszka, but other findings in this study pose a challenge to the authors' interpretation. These are the similar variable performance exhibited by Arab and Chinese participants at Low and Mid levels and the significant improvement of Chinese participants' production with rising proficiency. The RDH is a Full Transfer/ Partial Access account. This means that syntactic representations transfer from the L1 to L2 and UG fills any gaps unless where it is related to uninterpretable features. Past tense and verbal agreement are assumed to be associated with such features in the underlying syntax (see section 3.3.1). Accordingly, for the RDH while the Chinese participants here cannot instantiate these features, their Arab counterparts transfer them from their L1 from early on. However the performance of the two language groups at Low and Mid levels did not reflect this. Both language groups were similarly variable. In this sense, the source of variability might not be a permanent absence of features; this is because also Arab participants experienced variability on a par with Chinese participants and they did show that they acquired these features at the highest level of proficiency. Moreover, although the Chinese participants were variable at the High level diverging significantly from their Arab counterparts, they performed significantly better than their lower proficiency peers. It is difficult to account for the significant improvement with rising proficiency while assuming that their syntactic representations are absent permanently. Therefore, this performance is difficult to account for in the light of the RDH.

The structure building accounts of L2 acquisition (OG and MSBH; Vainikka and Young-Scholten, 1994, 1996, 1998, 2011, 2013; Hawkins, 2001) seem better equipped to explain the findings from the production study. Under these accounts, it is believed that the representations are absent initially and they project with rising proficiency. The developmental trajectories of both language groups across Low, Mid and High levels showed this clearly. These are presented in section 5.4.1 and 5.4.2 and repeated below in Figure 8.1, for convenience.

Figure 8.1: Trajectories of development with rising proficiency for Arab and Chinese participants (based on results presented in sections 5.4.1 and 5.4.2)



As shown in chapter five, the development with rising proficiency depicted in Figure 8.1 was statistically significant for the two language groups and in both past and verbal agreement. Given that the participants of this research are instructed learners and their performance show that they know the surface forms for past and agreement from the low level onward, it is very likely that the attested significant improvement is a reflection of a change in the underlying grammatical structure. This change seems to have taken place for Arab participants prior to their Chinese counterparts, whose performance indicates that they are in the process of this change. This demonstrates an L1 effect in this domain, a finding that favours the MSBH over OG.

One finding in the production study (and the perception study as well) still needs explanation. This is the great variation in the individual results attested in the two language groups, especially at Low and Mid levels. As can be seen above, the interpretation provided for the production study findings holds that the underlying representations for the properties under study were absent initially and they built up with rising proficiency. If this is correct, then it is very likely that the Low and Mid level participants used mainly their learned (classroom) knowledge to supply past and agreement on verbs. In turn, if the representations are really absent and learners use their learned knowledge to complete the task, it is not unexpected to encounter learners who differ in how they apply this knowledge in linguistic tasks. They might differ according to their views on how important they think it is to preserve these inflections in their production. However, even those who think it is very important would not be able to produce them consistently because they lack the underlying representations. Therefore, variation in individual results at the lower proficiency levels attested in this production study is consistent with the interpretation given to the variability in the production of the properties under study.

I turn to the results of perception study below

8.2.2 Results of the perception study

The results from the perception study (Chapter six) showed that the Arab and Chinese participants perceived the morphology similarly variably at the Low and Mid levels of proficiency, but at the High level, only the Chinese experienced this difficulty. Moreover, it was found that improvement with higher proficiency was observed only in the performance of Arab participants; the Chinese participants at the three proficiency levels performed similarly. Comparisons between Arab and Chinese participants at the three proficiency levels revealed no statistical differences, apart from an approaching significance difference between the two High level groups in their perception of verbal agreement (the advantage was for Arabic speakers). However, only Arab High level participants performed in a native-like manner, with their Chinese counterparts differing significantly from native speakers.

In pursuit of the source of perceptual limitations, on a par with the procedure followed in the production study, the results were broken down by allomorph, cluster and verb types. For convenience, the results of these analyses are summarised in Table 8.3.

Figure 8.2: A summary of the results of phonological analyses on past tense and verbal agreement morphology perception by all non-native groups

	Allomorphic variants (past)	Consonant cluster vs non-cluster (past)	Regular vs irregular (past)	Allomorphic variants (Agre)	Consonant cluster vs non-cluster (Agre)
Arab Low	All allomorphs variable No difference	No difference	No difference	All allomorphs variable No difference	No difference
Chinese Low	All allomorphs variable CV-d higher than VV-d	No difference	No difference	All allomorphs variable VV-z higher than CV-z, C-z and C-s; C-z higher than CV-z	No difference
Arab Mid	All allomorphs variable No difference	No difference	No difference	All allomorphs variable No difference	No difference
Chinese Mid	All allomorphs variable C-t higher than VV-d, CV-d and C-d	No difference	No difference	All allomorphs variable No difference	No difference
Arab High	All allomorphs accurate No differences	No difference	No difference	All allomorphs accurate No difference	No difference
Chinese High	All allomorphs variable No difference	No difference	No difference	All allomorphs variable No difference	No difference

Table 8.3 shows that the phonological effects are at the allomorph level. None of these effects branches from a constraint on coda consonant clusters as comparisons between consonant clusters and non-clusters show. This is neither due to a more general constraint on the perception of the regular morpheme as comparisons between regular and irregular verbs demonstrate.

The perception study revealed that the same participants who were found variable in their production of past tense and verbal agreement experienced perceptual difficulty with the same morphological items. The two language groups went similarly through a stage (Low and Mid levels) in which they had this difficulty. However, at a higher proficiency level (High level) only one language group (Arab) perceived the morphology accurately with the other (Chinese) being variable. This perceptual difficulty was not affected by the consonant cluster nature of the inflected verb coda or verb type (regular or irregular), which indicated that perceptual difficulty was not caused by poor phonological decoding abilities. This was not likely to be due to a lack of knowledge of the surface form either as these participants' production rates in the production study (chapter five) showed. Then, what causes the perceptual limitations?

One of the important findings in the literature discussed in Chapter three (section 3.2.2) is that some L2ers were found to have perceptual difficulties with inflectional morphology (e.g., Johnson and Newport, 1989; McDonald 2000). We assumed that these difficulties might reflect a syntactic problem. However, this interpretation was marred by findings of other studies showing that the perceptual difficulty might have been due to the poor phonological decoding abilities of L2ers (Solt et al., 2004; McDonald, 2006; McDonald and Roussel, 2010). These findings benefited this thesis study as they led us to control for the phonological shape of the verb to tease apart any effects that might occur in this domain. Yet, it was found that the participants in this research suffered from perceptual limitations but no phonological constraints on the perception of consonant clusters were detected. Therefore, what the perceptual limitations indicate in this case is that the source of L2ers' difficulties with the functional morphology might more likely be located deeper in the syntax.

Indeed, perceptual difficulties are expected to occur if the underlying representations are absent. This is inconsistent with the non-syntactic accounts of morphological variability discussed in this thesis (MSIH, PTH and PhRH). None of these accounts predicts the difficulty to extend from production to perception. This makes these accounts unsupported by not only the production but also the perception study results.

Therefore, the syntactic accounts - OG (Vainikka and Young-Scholten, 1994, 1996, 1998, 2011, 2013), MSBH (Hawkins, 2001) and RDH (Hawkins and Liszka, 2003) - are better candidates to explain the results from the perception study. The difference between the first two accounts (OG and MSBH) and the third one (RDH) is that only the latter attributes the problem to a permanent absence of the relevant syntactic representations from the Chinese learners' interlanguage grammars. The Chinese participants' overall results are consistent with the RDH as they experienced the perceptual difficulty at all proficiency levels. However, in the perception study (Chapter six, sections 6.4.1 and 6.4.2) it was found that some Chinese participants at the High level performed within the native speakers' range (4 out of 10 participants in past tense perception and 3 out of 10 participants in verbal agreement perception). These results are a challenge to the RDH as they might indicate that these participants have acquired the syntactic features. To verify this, I tracked down these participants' results in the processing data. The processing data confirmed that four participants performed similarly to native speakers (see appendix F for all individual results). These are participants Ch21 and Ch28 for past tense and Ch6 and Ch9 for verbal agreement. Their results are repeated here in Table 8.3 for convenience.

Table 8.3: Chinese participants who performed in a native-like manner

Participant NO	Perception %	RT mean	Proportion of target looks	First look to target mean
Ch21 (past)	93.33 %	0.974 sec	67%	-
Ch28 (past)	93.33 %	0.800 sec	62%	912 ms
Ch6 (Agreement)	100%	1.431 sec	58%	800 ms
Ch9 (Agreement)	91.66%	1.426 sec	64%	791 ms

sec= second; ms = millisecond

The native speakers' result range was presented in chapters six and seven for different types of data and repeated below in Table 8.4 for convenience:

Table 8.4: The native speakers' result range in different types of data

	Past	Agreement
Accuracy	80% to 100%	91.67% to 100%.
RT	0.774 to 1.443 sec	0.675 to 1.085 sec
First look to target	284 to 1081 ms	299 to 884 ms

sec= second; ms = millisecond

The proportion of eye looks to target show that these four Chinese participants looked reliably more to the target picture than the competitor. Their accuracy, RTs and speed of first look to target fell within the native speakers' results range (the RT in agreement is an exception).

Such performance is not expected by the RDH. It might be a sign of the presence of higher proficiency stage for the Chinese learners at which they acquire the syntactic features under study.

Again, the two structure building proposals (OG and MSBH) seem in a better position to account for the perception study results. The difficulty experienced by the two language groups at Low and Mid proficiency levels might be a reflection of absent syntactic representations. At the High level, while the performance of the Arab group signals that the underlying representations have been instantiated, the performance of their Chinese counterparts does not. However, as shown above, some Chinese participants performed within the native speakers' range. This suggests that Chinese learners might eventually acquire the syntactic features, but possibly later than their Arab counterparts. That the syntactic features were absent initially and projected with rising proficiency is compatible with both OG and MSBH. The apparent advantage for Arab participants at the High level gives more weight to the MSBH.

8.2.3 Results of the Processing Study

Results of the processing study (Chapter seven) came from three measures: 1) RT, 2) length of eye-look to target and 3) speed of first look to target (sensitivity measure). First, the RTs of the Arab and Chinese participants did not differ at any proficiency level (apart from RTs to agreement items where Arabs at the High level were faster at the approaching significance level). The two language groups exhibited significant progress (shorter RTs) with rising proficiency, but all groups were slower than their native counterparts. Secondly, the length of eye looks demonstrated robust differences between Arab and Chinese participants in past tense items as only the former at Mid and High proficiency levels looked reliably more to the target than competitor pictures and the latter showed no preference for either picture at any level. Eye movement patterns in verbal agreement trials, however, revealed similarities between the Arab and Chinese participants at the three proficiency levels as all looked reliably more to the target picture. Thirdly, the results of the sensitivity measure (the speed of convergence on target) exhibited no difference between the two language groups at Low and Mid levels, but a significant difference at the High level with Arabic speakers being faster. Results of this measure also revealed that only the Arab High level participants were as fast as native speakers and other groups were significantly slower.

As has been assumed in this thesis, given that the production, perception and processing use the same syntactic representations, the presence or absence of these should reflect in the three modalities similarly. The interpretation that has been given to the performance of research participants in production and perception studies is that the two language groups at Low and Mid levels lack the relevant syntactic representations. At the High level, however, while Arabs have them, the Chinese seem to be in the process of instantiating them. Therefore, if this interpretation is on the right track, it should reflect in the processing data.

How the presence or absence of representations would reflect in processing data is an issue discussed in Chapter seven (section 7.2.3). Following Jiang (2007), it was assumed that building up mental representations for specific properties is parallel to developing an automatic competence for them. Non-native participants were judged to have developed automatic competence based on their processing performance in the light of that of native speakers.

Indeed, consistent with the interpretation of production and perception results, the findings of the processing study demonstrated that only the Arab High-level group appeared to have processed the morphology as the group of native speakers did. This is the only group that is assumed to have the relevant representations. Although this group's RTs differed significantly from those of native speakers, the other processing data revealed that both groups were similar. Not only did the Arab High-level group look to target pictures reliably more as some other groups did as well, but also different from all other nonnative participants, their first look to the target picture (sensitivity measure) was as fast as that of native speakers. By contrast, all other non-native groups differed from native speakers in at least their RTS and speed of their first look to target.

Furthermore, the findings from the comparisons between the two language groups at the three proficiency levels in the processing study mirrored the picture that arose from the production and perception studies. The two language groups at Low and Mid levels processed the morphology similarly; their RTs, length of eye looks to target and speed of first look to target were similar (apart from length of eye looks to target in past tense items at Mid level and speed of first look in agreement items at Low level) and in all of these they performed significantly below native speakers. However, differences between the two language groups were more noticeable at the High level. In past tense items, only Arabic speakers looked more reliably to target pictures and they were significantly faster than their Chinese counterparts in their first look to target. In verbal agreement items, Arabic speakers were faster than the Chinese in their RTs (at the approaching significance level) and speed of first look to target (significantly).

All in all, the findings from the processing study confirm the interpretation given above to the production and perception results.

8.3 The Source of Adult L2 Learners' Morphological Variability

It was evident in the results that the variable use of past tense and verbal agreement morphology by the Chinese participants at the Low, Mid and High proficiency levels and Arab participants at the Low and Mid levels correlated with perceptual and processing limitations of the same morphological items. This indicated that variability was not merely due to the realization of surface forms; rather, the source was more likely to be a representational issue. As such, the MSIH (Prévost and White, 2000b) is not supported.

No phonological constraints at the prosodic level were detected in the production data. The detailed analysis of structures where the inflection is prosodified PWd internally and externally showed that variability affected them similarly. Chinese and Arabic allow the PWd internal structure but not the PWd adjunction (PWd external) and, hence, if a transfer at this level occurs, both language groups are expected to produce the inflection more consistently in the former than the latter structure. However, this was not attested in the data. No advantage for one structure over the other was observed. Therefore, the PTH (Goad, White and Steele, 2003; Goad and White, 2006) is not supported.

The results of the study also showed no phonological effects at the syllable level on the production or perception of past tense and verbal agreement morphology. The Arab participants at Low and Mid levels and Chinese participants at Low, Mid and High levels produced and perceived the regular past morpheme *-ed* and third-person *–s* across the four allomorphic variants inconsistently similarly with no preference to either of those forms that created word-final consonant clusters or those which did not (apart from the Arab Mid level participants who produced the *–ed* morpheme higher in non-consonant cluster contexts). Past tense on irregular verbs was also produced and perceived highly variably by the same groups. Accordingly, the PhRH (Lardiere, 1998a, 2003) is not supported.

The results indicate that variability is not caused by a permanent syntactic deficit. Although the performance of the Chinese High-level group did not show that they had acquired the properties under study, two particular findings indicated that it might not be a permanent state. The first finding is that significant improvement with rising proficiency was observed in the production and processing of the morphology. Given that these participants were instructed learners and their performance showed that they had metalinguistic knowledge at the lowest proficiency level, the significant improvement might be a reflection of a change in the relevant syntactic representations. The second finding is that some Chinese participants at the High level performed within the native speakers' range in the perception and processing of the morphology. This indicates that Chinese learners might be able to instantiate the required representations eventually. Therefore, the RDH (Hawkins and Liszka, 2003) is not supported.

The study findings revealed that the source of variability is more likely located deeper at the syntax. It was observed that the two language groups at Low and Mid proficiency levels similarly experienced noticeable limitations in production, perception and processing of past and verbal agreement. This indicates that the syntactic representations might be absent at this stage. Then, significantly better performance was seen in higher proficiency Arab and Chinese groups. This development with rising proficiency seems to mirror a gradual build-up of syntactic representations. Taken together, the results are consistent with the structure building accounts (i.e., OG (Vainikka and Young-Scholten, 1994, 1996, 1998, 2011, 2013) and MSBH (Hawkins, 2001)).

The L2ers' native language is found to play a role in the acquisition of the linguistic properties under study. An L1 effect was not seen at lower proficiency levels as both Arab and Chinese Low and Mid level groups' performance was strikingly similar, but the two language groups diverged at the High level with the Arabic speakers outperforming their Chinese counterparts. With the absence of phonological effects on the data, the source of the observed difference between the performance of the two language groups is very likely to be L1 syntax. Recall that only Arabic marks verbs for past and agreement and thus it has been assumed that while Arabic has the relevant syntactic features, Chinese does not. This L1 difference did not show up in the performance of the Low and Mid level groups, which is consistent with the claim that the relevant syntactic representations were absent from the interlangauage grammars of these learners. The L1 effect, however, arose at the High proficiency level where the two language groups performed significantly better than their lower proficiency peers. This might suggest that the effect took place as the syntactic representations were being instantiated. Therefore, this is fully compatible with the MSBH (Hawkins, 2001)

Returning to the research question "*what is the source of morphological variability and its persistence in the use of past tense and verbal agreement morphology in adult SLA of English?,* we can conclude with the following answer based on the findings of the production, perception and processing studies. The source of variability in the use of past tense and verbal agreement seems to be a temporary absence of the relevant syntactic representations. In turn, the absence of these syntactic features from the L1 is likely to be what makes learners take longer to instantiate the L2 features. Finally, the findings indicate that these representations build up with rising proficiency and thus the variability can be overcome.

8.4 Limitation of the Studies and Future Research

One limitation of the present study is that it did not control for the frequency of verbs included in the stimuli of the elicited imitation and picture choice tasks. I have argued that the variability in the production and perception of morphology by the Low and Mid level learners is caused by the absence of the corresponding underlying representations and that L2ers rely mainly on their learned knowledge to perform linguistic tasks. If this is truly the case, how frequent the verb is in the linguistic input might have an effect on its production or perception. Therefore, this aspect should be taken into consideration in future research.

Another aspect in the design of this study to be improved by future research is the participant population. Testing untutored L2 learners is a promising research perspective in the investigation of morphological variability. The present research has investigated the issue in instructed learners. The grammatical properties under study here are taught early in classrooms. This might have an effect on the performance of the participants in this research. Therefore, the study of untutored learners would eliminate this effect and, therefore, lead to a better assessment of the phenomenon.

The study of the perception and processing of inflectional morphology is another promising research perspective as it provides new insights into the source of morphological variability. The present research used a picture choice task supported by RT and eye tracking techniques in this endeavour. Although this attempt appears to be successful, a replication is needed to validate the findings of this research.

Furthermore, to locate the source of morphological variability, the research at hand has examined past tense and verbal agreement morphology in L2 English. However, the phenomenon of morphological variability is also attested in other grammatical properties in English as well as other languages. Therefore, further research is needed to examine whether the same results would be obtained upon examining the perception and processing of other grammatical properties in L2 English or other languages.

8.5 Conclusion

This thesis is an attempt to locate the source of morphological variability in the use of past tense and verbal agreement morphology and the reason for its persistence in some L2ers rather than others. Different from previous research on the phenomenon, this research examined not only the production, but also the perception and processing of the properties under study. It also differed from previous endeavours as it used a cross-sectional design which included learners at different points of linguistic development (i.e., Low, Mid and High proficiency levels). By doing this, a broader view of the underlying representations giving rise to the phenomenon was obtained. This also provided testing grounds for six of the prominent accounts advanced in the field.

The role of the L1 was also explored in the research at hand. L2ers of English from two different L1 backgrounds (Arabic and Mandarin Chinese) were included. The similarities and dissimilarities between the two languages, on one hand, and between each of them and the target L2, on the other, allowed for testing a number of proposals on the role of the L1 in morphological variability.

The results of the studies showed no phonological effects at the syllable or prosodic levels, which disconfirmed the Phonological Reduction Hypothesis (Lardiere, 1998a and 2003) and the Prosodic Transfer Hypothesis (Goad, White and Steele, 2003; Goad and White, 2006). The Missing Surface Inflection Hypothesis (Prévost and White, 2000a, b), which attributed morphological variability to processing failure (during production) between target like representations and surface forms, was not supported either as learners' difficulty with the morphology was found to extend from production to perception and processing. As some of the Chinese participants performed within the native speaker range in the perception task, the proposal that their problems were caused by a permanent syntactic deficit (i.e., Representational Deficit Hypothesis, (Hawkins and Liszka, 2003)) was also not supported. The structure building accounts (i.e., Organic Grammar (Vainikka and Young-Scholten, 1994, 1996, 1998, 2011, 2013) and Modulated Structure Building Hypothesis (Hawkins, 2001)) were found better equipped to explain the observed patterns in production, perception and processing. The findings that the difficulty in production correlated with difficulties in the perception and processing and both language groups performed significantly better with rising proficiency can be better accounted for under the light of structure building accounts of SLA.

References

Abu-Mansour, M. (1991) 'Epenthesis in Makkan Arabic: Unsyllabifiable consonants versus degenerate syllables' in Eid, M. and McCarthy, J. (eds.) *Perspectives on Arabic Linguistics III: Papers from the third annual symposium on Arabic linguistics*. Amsterdam: John Benjamins. pp. 137-159.

Abney, S. (1987) *The English noun phrase in its sentential aspect*. Doctoral Dissertation, MIT.

Adamson, H. D. (1988) *Variation Theory and Second Language Acquisition*. Washington DC: Georgetown University Press. (cited in Unsworth, 2008)

Adger, D. (2003) *Core Syntax: A Minimalist Approach*. Oxford; New York: Oxford University Press.

Al-Aqarbeh, R. (2011) *Finiteness in Jordanian Arabic: A Semantic and Morphosyntactic Approach*. Dissertation. University of Kansas. URI: http://hdl.handle.net/1808/9820

Al-Azraqi, M. (1998) *Aspects of the Syntax of the Dialect of Abha (South West Saudi Arabia)*. PhD Thesis. University of Durham, Durham, UK.

Allan, D. (1992) *Oxford Placement Test*. Oxford: Oxford University Press.

Anderson, J. I. (1978) 'Order of difficulty in adult second language acquisition', in Ritchie, W. C (ed.). *Second Language Acquisition Research: Issues and Implications.* New York: Academic Press. pp. 91-108.

Antony, L. and Hornstein, N. (eds.) (2002) *Chomsky and His Critics*. Oxford: Blackwell.

Aoun, E. J., Benmamoun, E. & Choueiri, L. (2010) *The Syntax of Arabic*. Cambridge: Cambridge University Press.

Aquil, R., (2013) 'Cairene Arabic syllable structure through different phonological theories', *Open Journal of Modern Linguistics (OJML)*, 3(3), pp. 259-267.

Arnold, J. E., Eisenband, J. G., Brown-Schmidt, S., & Trueswell, J. C. (2000) 'The rapid use of gender information: Evidence of the time course of pronoun resolution from eye tracking', *Cognition*, 76, pp. 13-26.

Bailey, N. Madden, C. & Krashen, S.D. (1974), 'Is there a 'Natural Sequence' in adult second language learning?' *Language learning,* 24(2), pp. 235-243.

Barto-Sisamout, K., Nicol, J., Witzel, J., & Witzel, N. (2009) 'Transfer effects in bilingual sentence processing', *Arizona Working Papers in Second Language Acquisition and Teaching,* 16*,* pp. 1-26.

Bayley, R. J. (1991), *Variation Theory and Second Language Learning: Linguistics and Social Constraints on Interlanguage Tense Marking*, PhD dissertation. Stanford University.

Bayley, R. J. (1996), 'Competing constraints on variation in the speech of adult Chinese learners of English', in Bayley, R. J. & Preston, D. (eds.), *Second Language Acquisition and Linguistic Variation*. Amsterdam: John Benjamins.

Beck, M.L. (1998) 'L2 Acquisition and Obligatory Head Movement: English Speaking Learners of German and the Local Impairment Hypothesis', *Studies in Second Language Acquisition* 20, pp. 311-348.

Benmamoun, E. (2000) *The Feature Structure of Functional Categories: A Comparative Study of Arabic Dialects*, New York, Oxford: Oxford University Press.

Bley-Vroman R. and Chaudron, C. (1994) 'Elicited imitation as a measure of second-language competence', in Tarone, E.E., Gass, S. and Cohen, A.D. (eds), *Research Methodology in Second Language Acquisition*. Hilldale: Lawrence Erlbaum, pp. 245–261

Bond, K., Gabriele, A., Fiorentino, R. and Aleman Bañón, J. (2011) 'Individual Differences and the Role of the L1 in L2 Processing: An ERP Investigation', in Herschensohn, J. & Tanner, D. (Eds.) *Proceedings of the 11th Generative Approaches to Second Language Acquisition Conference (GASLA).*Somerville, MA: Cascadilla Proceedings Project. pp. 17-29

Bosisio, N. (2003) *Vocoids and Their prosodic Distribution, with Special Reference to Italian and Arabic*. Durham theses, Durham University. Available at Durham E-Theses Online: http://etheses.dur.ac.uk/4076/

Btoosh, M. (2006) 'Constraint Interactions in Jordanian Arabic Phonotactics: An Optimality-Theoretic Approach'. *Journal of Language and Linguistics*. 5(2), pp. 192-221

Campos Dintrans, G. (2011) *Acquisition of morphosyntax in the adult second language: the phonology factor.* Dissertation, University of Iowa. URL: http://ir.uiowa.edu/etd/2677.

Chen, L., Shu, H., Liu, Y., Zhao, J. & Li, P. (2007) 'ERP signatures of subject–verb agreement in L2 learning', *Bilingualism: Language and Cognition,* 10*,* pp. 161–174.

Chomsky, N. (1959) 'A Review of B. F. Skinner's Verbal Behavior', *Language*, 35(1), pp. 26- 58.

Chomsky, N. (1965) *Aspects of the Theory of Syntax*. MIT Press.

Chomsky, N. (1972) *Language and Mind*. New York: Harcourt, Brace & Jovanovich.

Chomsky, N. (1981) *Lectures on Government and Binding: the Pisa lectures*. Dordrecht: Foris.

Chomsky, Noam (1986) *Barriers*. Cambridge, MA: MIT Press.

Chomsky, N. (1991) *The Sound Pattern of English*. Cambridge, Mass.: MIT Press.

Chomsky, N. (1993) 'A minimalist program for linguistic theory'. in Hale, K. & Keyser, S.J. (eds.). *The view from building 20: Essays in linguistics in honor of Sylvain Bromberger,* Cambridge, MA: MIT Press. pp. 1-52.

Chomsky, N. (1995) *The Minimalist Program*. MIT Press.

Chomsky, N. (2000). 'Minimalist inquiries: The framework', in Martin, R., Michaels, D. and Uriagereka, J. (eds.) *Step by Step: Essays in Minimalist Syntax in Honor of Howard Lasnik.* Cambridge, MA: MIT Press. pp. 89-155.

Chomsky, N. (2001) 'Derivation by Phase', in Kenstowicz, M. (ed.), *Ken Hale: a Life in Language*. MIT press

Clahsen, H. (2007) 'Psycholinguistic perspectives on grammatical representations', in Featherston S. & Sternefeld, W. (eds.) *Roots: Linguistics in search of its evidential base*. Berlin: Mouton de Gruyter. pp.97-132.

Clahsen, H. and Felser, C. (2006) 'Grammatical processing in language learners', *Applied Psycholinguistics,* 27, pp. 3-42.

Clahsen, H., Felser, C., Neubauer, K., Sato, M. & Silva, R. (2010) 'Morphological structure in native and non-native language processing', *Language Learning,* 60, pp. 21-43.

Cooper, R. M. (1974) 'The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing', *Cognitive Psychology*, 6, pp. 84–107.

Cowell, M. W. (1964) *A reference grammar of Syrian Arabic, based on the dialect of Damascus*. Washington, DC: Georgetown University Press.

Dulay, H.C. & Burt, M.K. (1973) 'Should we teach children syntax?', *Language Learning,* 23, pp. 245-258.

Dulay, H.C. & Burt, M.K. (1974) 'Natural sequences in child second language acquisition', *Language Learning,* 24, pp. 37-53.

Dussias, P. E. and Pinar, P. (2009) 'Sentence Parsing in L2 Learners: Linguistic and Experience-Based Factors', in Ritchie, W. C. and Bhatia, T. K. *The New Handbook of Second Language Acquisition* (2nd edition). Bingley, UK: Emerald Group Publishing Limited. pp. 296-318.

Duanmu, S. (2000) *The phonology of Standard Chinese*. Oxford University Press. (cited in Goad and White, 2006)

Ellis, R. & Barkhuizen, G. (2005) *Analysing learner language*. Oxford: Oxford University Press.

Epstein, S.D., Flynn, S. & Martohardjono, G. (1996) 'Second language acquisition: Theoretical and experimental issues in contemporary research', *Brain Behavioural Sciences,* 19(4), pp. 677-714.

Epstein, S.D., Flynn, S. & Martohardjono, G. (1998) 'The strong continuity hypothesis: Some evidence concerning functional categories in adult L2 acquisition', in Flynn, S., Martohardjono, G. & O'Neil, W. (eds.), *The Generative Study of Second Language Acquisition.* Lawrence Erlbaum Assoc Inc, pp. 61-77.

Eubank, L. (1993/1994) 'On the transfer of parametric values in L2 development', *Language Acquisition, 3, pp. 182-208.*

Eubank, L., Bischof, J., Huffstutler, A., Leek, P. & West, C. (1997) '"Tom eats slowly cooked eggs": Thematic-verb raising in L2 knowledge', *Language Acquisition* 6, pp. 171-99.

Erlam, R. (2006) 'Elicited imitation as a measure of L2 implicit knowledge: An empirical validation study', *Applied Linguistics,* 27, pp. 464-491.

Erlam, R. (2009) 'Elicited imitation as a measure of L2 implicit knowledge', in Ellis, R., Loewen, S., Elder, C., Erlam, R., Philp, J. & Reinders, H. (Eds.) *Implicit and Explicit Knowledge in Second Language Learning, Testing and Teaching*. Bristol: Multilingual Matters. pp. 65-93.

Erwin, W. M. (1963) *A short reference grammar of Iraqi Arabic*. Washington: Georgetown University Press.

Foley C. and Flynn, S. (2013) 'The role of the native language', in Herschensohn, J. and Young-Scholten, M. *The Cambridge Handbook of Second Language Acquisition.* Cambridge: Cambridge University Press. pp.97-113

Frenck-Mestre, C. (2002) 'An on-line look at sentence processing in a second language', in Herrida, R. and Altarriba, J. (Eds.) *Bilingual Sentence Processing*. Elsevier. pp.217-236

Frenck-Mestre, C., Foucart, A., Carrasco, H., & Herschensohn, J. (2009) 'Processing of grammatical gender in French as a first and second language: Evidence from ERPs', in Roberts, L., Véronique, D., Nilsson, A. and Tellier, M. (Eds.) *EUROSLA Yearbook*. Benjamins. pp 76-106.

Franceschina, F. (2001) 'Morphological or syntactic deficits in near-native speakers? An assessment of some current proposals', *Second Language Research,* 17(3), pp. 213-247.

Gass, S. M, and Selinker, L. (2008) *Second Language Acquisition: An Introductory Course* (3rd ed). New York, NY: Routledge.

Goad, H. and White, L. (2006) 'Ultimate attainment in interlanguage grammars: A prosodic account', *Second Language Research,* 22, pp. 243-268.

Goad, H., White, L. & Steele, J. (2003) 'Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations?', *The Canadian Journal of Linguistics/La Revue Canadienne De Linguistique*, *48*(3-4), pp. 243-263.

Granena, G., & Long, M.H. (2013) 'Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains, *Second Language Research*, 29, pp. 311-343.

Grondin, N. & White, L. (1996) 'Functional categories in child L2 acquisition of French', *Language Acquisition,* 5, pp. 1-34.

Guasti, M.T. (2002) *Language Acquisition: The growth of grammar*. Cambridge, MA: MIT Press.

Guiraud, P. (1960) *Problemes en methodes de la statistique linguistique*. Dordrecht: D. Reidel (cited in Unsworth 2008)

Halle, M. & Marantz, A. (1993) 'Distributed morphology and the pieces of inflection', in Hale, K. & Keyser, S.J. (eds.) *The view from building 20: Essays in linguistic honour of Sylvain Bromberger*. Cambridge, Mass: MIT Press. pp. 111-176

Hamayan, E., Saegert, J. & Laraudee P. (1977) 'Elicited imitation in second language learners', *Language and Speech,* 20, pp. 86–97.

Hansen, J. (2001) 'Linguistic constraints on the acquisition of English syllable codas by native speakers of mandarin Chinese', *Applied Linguistics* 22, pp. 338-365.

Hawkins, R. (2001) *Second Language Syntax: A Generative Introduction*. Oxford: Blackwell.

Hawkins, R. (2009) 'Second language acquisition of morphosyntax', in Ritchie, W. C. & Bhatia, T. K. (eds.) *The New Handbook of Second Language Acquisition* (2nd edition). Bingley, UK: Emerald Group Publishing Limited. pp. 211-236

Hawkins, R. and Chan, C. (1997) 'The partial availability of Universal Grammar in second language acquisition: The 'failed functional feature hypothesis', *Second Language Research,* 13, pp. 187-226.

Hawkins, R., & Hattori, H. (2006) 'Interpretation of English multiple wh-questions by Japanese speakers: a missing uninterpretable feature account', *Second Language Research, 22*, pp. 269-301.

Hawkins, R., & Liszka, S. (2003) 'Locating the source of defective past tense marking in advanced L2 English speakers', in Van Hout, R., Hulk, A., Kuiken, F. & Towell, R. (eds.), *The lexicon-syntax interface in second language acquisition.* Amsterdam; Netherlands: John Benjamins. pp. 21-44.

Haznedar, B., & Schwartz, B. D. (1997) 'Are there optional infinitives in child L2 acquisition?', in Hughes, E. and Greenhill, A. (eds.) *Proceedings of the 21st Annual Boston University Conference on Language Development.* Sommerville, MA: Cascadilla Press. pp. 257-268.

Hayhoe, M., & Ballard, D. (2005) 'Eye movements in natural behavior' *Trends in Cognitive Sciences*, 9, pp. 188–194.

Herschensohn, J. (2001) "Missing Inflection in Second Language French: Accidental Infinitives and Other Verbal Deficits", *Second Language Research,* 17, pp.273-305.

Hornstein, N., Nunes, J. and Grohmann, K. K. (2005) *Understanding Minimalism*. Cambridge: Cambridge University Press.

Hornstein, N. & Lightfoot, D. (1981) 'Introduction', in Hornstein, N. & Lightfoot, D (ed.) *Explanation in Linguistics: The Logical Problem of Language Acquisition*. London: Longman. pp. 9-31.

Hulstijn, J. H., & Hulstijn, W. (1984) 'Grammatical errors as a function of processing constraints and explicit knowledge', *Language Learning,* 34, pp. 23-43. (cited in Erlam, 2009)

Hunt, K. W. (1970) *Syntactic Maturity in Schoolchildren and Adults*. Chicago IL: University of Chicago Press. (cited in Unsworth, 2008)

Ionin, T. (2013) 'Morphosyntax', in Herschensohn, J. and Young-Scholten, M. *The Cambridge Handbook of Second Language Acquisition.* Cambridge: Cambridge University Press. pp. 505-528.

Ionin, T. & Wexler, K. (2002) ‘Why is ‘is’ easier than ‘-s’: Acquisition of tense/ agreement morphology by child second language learners of English’, *Second Language Research,* 18(2), pp.95-136.

Ito, J. and Mester, A. (1992) ‘Weak Layering and Word Binarity’*, Linguistics Research Center Report Santa Cruz: University of California*. pp. 92-09. (cited in Kiprasky, 2003)

Jiang, N. (2004) ‘Morphological insensitivity in second language processing.’ *Applied Psycholinguistics,* 25, pp. 603-634.

Jiang, N. (2007) ‘Selective integration of linguistic knowledge in adult second language acquisition’ *Language Learning,* 57, pp.1-33.

Johnson, J. S. (1992) ‘Critical period effects in second language acquisition: The effect of written versus auditory materials on the assessment of grammatical competence’, *Language Learning,*42, pp. 217–248. (cited in Jiang 2004)

Johnson, J.S. & Newport, E.L. (1989) ‘Critical period effects in second language learning: The influence of maturational state on the acquisition of English as a second language’, *Cognitive Psychology,* 21, pp. 60-99.

Kayne, R. S. (1984) *Connectedness and Binary Branching*. Dordrecht: Foris.

Keating, Gregory D. (2009) ‘Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation’, *Language Learning,* 59, pp. 503-35.

Kempe, V. and MacWhinney, B. (1998) ‘The acquisition of case marking by adult learners of Russian and German’, *Studies in Second Language Acquisition,* 20, pp. 543–87.

Kiprasky, P. (2003) ‘Syllables and moras in Arabic’, in Fery, C. & Vijver, R. V. D. (Eds.) *The syllable in optimality theory*. New York, NY: Cambridge University Press.

Krashen, S. D. (1977) ‘The monitor model for adult second language performance’, in Burt, M., Dulay, H. and Finocchiaoro, M. (eds.) *Viewpoints on English as a second language*. New York: Regents. pp. 152–161.

Lado, R. (1957) *Linguistics across cultures*. Ann Arbor: University of Michigan

Laradi, W.J. (1983) *Pharyngealization in Libyan (Tripoli) Arabic: An instrumental study*. PhD thesis, University of Edinburgh. URL: http://hdl.handle.net/1842/6717

Lardiere, D. (1998a) ‘Case and tense in thefossilized steady state’, *Second Language Research,* 14, pp. 1-26.

Lardiere, D. (1998b) 'Dissociating syntax from morphology in a divergent L2 end-state grammar', *Second Language Research,* 14(4), pp. 359-375.

Lardiere, D. (2000) 'Mapping features to form in second language acquisition', in Archibald, J (ed.) *Second language acquisition and linguistic theor*y. Oxford: Blackwell. pp. 102-29.

Lardiere, D. (2003) 'Second Language Knowledge of [+/–Past] vs. [+/–Finite]', in M. Liceras, J. M., Zobl, H. ad Goodluck, H. (eds.) *Proceedings of the 6th Generative Approaches to Second Language Acquisition Conference (GASLA 2002)*. Somerville, MA: Cascadilla Proceedings Project. pp. 176-189.

Lardiere, D. (2007) *Ultimate attainment in second language acquisition: a case study*. Mahwah, NJ: Lawrence Erlbaum Associates.

Lardiere, D. (2008) 'Feature assembly in second language acquisition' in Liceras, J.M., Zobl, H. and Goodluck, H. (eds) *The role of formal features in second language acquisition*. New York: Lawrence Erlbaum Associates. pp. 106–40.

Lardiere, D. (2009) 'Some thoughts on a contrastive analysis of features in second language acquisition', *Second Language Research, 25*(2), pp. 171–225.

Lardiere, D. (2013) 'Nativelike and non-nativelike attainment', in Herschensohn, J. and Young-Scholten, M. *The Cambridge Handbook of Second Language Acquisition.* Cambridge: Cambridge University Press. pp.670-692

Larsen-Freeman, D. (1975) 'The acquisition of grammatical morphemes by adult ESL students', *TESOL Quarterly,*9, pp. 409–30.

Larsen-Freeman, D. E. and Strom, V. (1977) 'The construction of a second language acquisition index of development' *Language Learning,* 27, pp.123–134. (cited in Unsworth, 2008)

Larson-Hall, J. (2010) *A Guide to Doing Statistics in Second Language Research Using SPSS*. New York and London: Routledge

Levy, J. (1999) *Totally Uncool*. Illustrated by Chris Monroe. Minneapolis: Carolrhoda Books.

Li, Ch. and Thompson, S. (1989) *Mandarin Chinese: A Functional Reference Grammar.* Berkeley and Los Angeles: University of California Press.

Liversedge, S. P., & Findlay, J. M. (2000) 'Saccadic eye movements and cognition', *Trends in Cognitive Sciences*, 4, pp. 6–14.

Mackey, A. & Gass. S. (2005) *Second language research: Methodology and design*. Mahwah, NJ: Lawrence Erlbaum.

Makino, T. (1980) 'Acquisition order of English morphemes by Japanese secondary school students', *Journal of Hokkaido University of Education,* 30, pp.101-148.

Mahfoudhi, A. (2005) 'Moraic Syllable Structure and Edge Effects in Arabic', in Alhawary, M. T. & Benmamoun, E. (eds.) *Perspectives on Arabic Linguistics: Papers from the annual symposium on Arabic linguistics. Volume XVII–XVIII*. Benjamins. pp. 27–48.

McDade, H.L., Simpson, M.A. & Lamb, D.E. (1982) 'The use of elicited imitation as a measure of expressive grammar: a question of validity', *Journal of Speech and Hearing Disorders,* 47, pp. 19–24.

McDonald, J. (2000) 'Grammaticality judgments in a second language: Influences of age of acquisition and native language', *Applied Psycholinguistics,* 21, pp. 395-423.

McDonald, J. (2006) 'Beyond the critical period: Processing-based explanations for poor grammaticality judgment performance by late second language learners', *Journal of Memory and Language,* 55, pp. 381-401.

McDonald, J. & Roussel, C. (2010) 'Past tense grammaticality judgment and production in non-native and stressed native English speakers', *Bilingualism: Language and Cognition,* 13, pp.429–448.

Meisel, J. (1991) 'Principles of UG and strategies of language use: on some similarities and differences between first and second language acquisition', in Eubank, L. (ed.) *Point-Counterpoint*. Amsterdam: Kluwer, 231-276.

Meisel, J.M. (1997) 'The acquisition of the syntax of negation in French and German: Contrasting first and second language development', *Second Language Research,* 13, pp. 227-263.

Murphy, G. L., & Shapiro, A. M. (1994) 'Forgetting of verbatim information in discourse', *Memory & Cognition*, 22, pp. 85-94. (cited in Erlam, 2009)

Nation, I. P. S. (1990) *Teaching and Learning Vocabulary*. Boston, MA: Heinle and Heinle.

Nespor, M. and Vogel, I. (1986) *Prosodic Phonology*. Dordrecht: Foris.

Ojima, Shiro, Nakata, Hiroki, & Kakigi, Ryusuke. (2005) 'An ERP study of second language learning after childhood: effects of proficiency', *Journal of Cognitive Neuroscience,* 17, pp. 1212-1228.

Oldenkamp, L. (2013) *The trouble with inflection for adult learners of Dutch: A study of the L1-L2 interplay of morphosyntactic and phonetic-phonological factors*. Utrecht, the Netherlands: LOT.

Ortega, L. (2003) 'Syntactic complexity measures and their relationship to L2 proficiency: A research synthesis of college-level L2 writing', *Applied Linguistics* 24, pp. 492–518. (cited in Unsworth, 2008)

Owens, J (1984) *A Short Reference Grammar of Eastern Libyan Arabic*, Wiesbaden: Harrassowitz.

Pérez-Leroux, A. T. & Glass, W. R. (1997) 'OPC effects on the L2 acquisition of Spanish', in Pérez-Leroux, A. T. & Glass, W. R. (eds.) *Contemporary perspectives on the acquisition of Spanish*, Vol. 1: *Developing grammars*. Sommerville MA: Cascadilla Press. pp. 149–165.

Pinker, S., & Ullman, M. T. (2002) 'The past and future of the past tense', *Trends in Cognitive Sciences, 6*(11), pp. 456-463.

Platzack, C. (1996) 'The initial hypothesis of syntax: A minimalist perspective on language acquisition and attrition' in Clahsen, H. (ed.) *Generative perspectives on language acquisition*. Amsterdam: John Benjamins. pp. 369-414.

Po-Ching, Yip and Rimmington, Don (2006; 2nd ed.). *Chinese: An Essential Grammar*. Oxford: Taylor & Francis (Routledge).

Pollock, J-Y. (1989) 'Verb Movement, Universal Grammar, and the Structure of IP', *Linguistic Inquiry* 20, pp. 365-424.

Prévost, P. (2003) 'The Issue of Morphological Variation in Adult L2 French', in Van Kampen, J. and Baauw, S. (eds) *Proceedings of GALA 2003 (Generative Approaches to Language Acquisition)*. Utrecht, The Netherlands: LOT.

Prévost, P. and L. White (2000a) 'Accounting for morphological variability in second language acquisition: Truncation or Missing Inflection?', in Friedemann M.-A. and Rizzi L. (eds.) *The Acquisition of Syntax*. London: Longman. pp. 202-235.

Prévost, P. and White, L. (2000b) 'Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement', *Second Language Research,* 16, pp. 103-133.

Prince, A. and Smolensky, P. (1993) *Optimality Theory: Constraint interaction in generative grammar.* Boulder: Rutgers University and University of Colorado.

Radford, A. (2004) *Minimalist Syntax: Exploring the Structure of English.* Cambridge: Cambridge University Press.

Richards, B. (1987) 'Type/token rations: What do they really tell us?', *Journal of Child Language,*14, pp. 201–209. (cited in Unsworth, 2008)

Roberts, L & Siyanova-Chanturia, A 2013, 'Using eye-tracking to investigate topics in L2 acquisition and L2 sentence/discourse processing', *Studies in Second Language Acquisition*, 35(2), pp. 213-235.

Rossi, S., Gugler, M., Friederici, A. and Hahne, A. (2006) ‘The impact of proficiency on syntactic second-language processing of German and Italian: Evidence from event related potentials’, *Journal of Cognitive Neuroscience,* 18, pp. 2030-2048

Ryding, K. C. (2005) *A Reference Grammar of Modern Standard Arabic*. Cambridge University Press.

Sato, M. & Felser, C. (2008) ‘Sensitivity to morphosyntactic violations in English as a second language’, Ms. University of Essex.

Schwartz, B. D. (1998) ‘The second language instinct’, *Lingua,* 106(1-4), pp. 133-160.

Schwartz, B.D. & Sprouse, R.A. (1994) ‘Word order and Nominative Case in non native language acquisition: A longitudinal study of L1 Turkish German interlanguage’, in T. Hoekstra, T. & Schwartz, B.D. (eds.) *Language Acquisition Studies in Generative Grammar*. Amsterdam: John Benjamins. pp. 317-368.

Schwartz, B. & Sprouse, R. (1996) ‘L2 cognitive state and the Full Transfer/Full Access Model’, *Second Language Research*, 12, pp. 40-72.

Selkirk, E. O. (1986) ‘On derived domains in sentence phonology’, *Phonology,* 3, pp. 371–405.

Selkirk, E. O. (1997) ‘The prosodic structure of function words’, in Morgan, J. and Demuth, K. (eds) *Signal to syntax: Bootstrapping from speech to grammar in early acquisition.* Mahwah: Lawrence Erlbaum. pp. 187–213.

Solt, S., Pugach, Y., Klein, E. C., Adams, K., Stoyneshka, I., & Rose, T. (2004) ‘L2 perception and production of the English regular past: Evidence of phonological effects’, in Brugos, A., Micciulla, L. & Smith, C. E. (Eds.), *BUCLD 28: Proceedings of the 28th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press. pp. 553–564.

Stauble, A.M. (1984) ‘A comparison of a Spanish-English and a Japanese- English second language continuum: Negation and verb morphology’ in Andersen, R. (ed.) *Second Language: A cross-linguistic perspective.* Newbury. pp. 323-353.

Tatham, M. & Morton, K. (2006) *Speech Production and Perception.* New York: Palgrave Macmillan.

Tatsuno, Y. and Sakai K.L. (2005) 'Language-related activations in the left prefrontal regions are differentially modulated by age, proficiency, and task demands', *Journal of Neuroscience*, 25, pp. 1637-1644.

Tanenhaus, M. & Trueswell, J. (2006) 'Eye Movements and Spoken Language Comprehension', in Traxler, M. and Gernsbacher, M. (Eds.) *The Handbook of Psycholinguistics.* Elsevier (2nd edition). pp. 863-900.

Tsimpli, I. M. (2003) 'Interrogatives in the Greek/English interlanguage: a minimalist account', in Mela-Athanasopoulou, E. (ed.) *Selected papers on theoretical and applied linguistics*. Thessaloniki: Aristotle University. pp. 214-225.

Tsimpli, I.M., & Dimitrakopoulou, M. (2007) 'The interpretability hypothesis: Evidence from wh-interrogatives in second language acquisition', *Second Language Research* 23(2), pp. 215-242.

Tsimpli, I. M., & Roussou, A. (1991) 'Parameter resetting in L2?', *UCL Working Paper in Linguistics* 3, pp. 149-169.

Unsworth, S. (2005) *Child L2, Adult L2, Child L1: Differences and Similarities. A study on the acquisition of direct object scrambling in Dutch*. PhD thesis, Utrecht University.

Unsworth, S. (2008) 'Comparing Child L2 Development with Adult L2 Development: How to Measure L2 Proficiency' in Haznedar, B. and Gavruseva, E. (eds.) *Current Trends in Child Second Language Acquisition.* Amsterdam: Benjamins. pp. 301-333.

Vainikka, A. & Young-Scholten, M. (1994) 'Direct access to X-bar theory: Evidence from Korean and Turkish adults learning German', in Hoekstra, T. & Schwartz, B.D. (eds.) *Language Acquisition Studies in Generative Grammar*. Amsterdam: Benjamins. pp. 265-316.

Vainikka, A. & Young-Scholten, M. (1996a) 'Gradual development of L2 phrase structure', *Second Language Research,* 12(1), pp.7-39.

Vainikka, A. & Young-Scholten, M. (1996b) 'The early stages in adult L2 syntax: Additional evidence from Romance speakers', *Second Language Research,* 12(2), pp. 140-176.

Vainikka, A. & Young-Scholten, M. (1998) 'Morphosyntactic triggers in adult SLA', in Beck, M.L. (ed.) *Morphology and its interface in second language knowledge*. Amsterdam: Benjamins. pp. 89-113.

Vainikka, A. & Young-Scholten, M. (2005) ‘The roots of syntax and how they grow Organic Grammar, the Basic Variety and Processability Theory’, in Unsworth, S., Sorace, A., Parodi, T. and Young-Scholten, M. (eds) *Paths of Development in L1 and L2 Acquisition*. Amsterdam: Benjamins.

Vainikka, A. and Young-Scholten, M. (2007) ‘Minimalism vs. Organic Syntax', in Karimi, S., Samiian, V. and Wilkins, W.K. (eds.) *Clausal and Phrasal Architecture: Syntactic Derivation and Interpretation*. Amsterdam: Benjamins, pp. 319-338.

Vainikka, A. and Young-Scholten, M. (2011) *The Acquisition of German, Introducing Organic Grammar*. Berlin: Walter de Gruyter GmbH & Co. KG.

Vainikka, A. and Young-Scholten, M. (2013) ‘Stagelike development and Organic Grammar’, in Herschensohn, J. and Young-Scholten, M. *The Cambridge Handbook of Second Language Acquisition.* Cambridge: Cambridge University Press. pp.581-604

Vinther, T. (2002) ‘Elicited imitation: a brief review’, *International Journal of Applied Linguistics*, 12, pp. 54–73.

Wartenburger, I., Heekeren, H., Abutalebi, J., Cappa, S., Villringer, A. and Perani, D. (2003) ‘Early setting of grammatical processing in the bilingual brain’, *Neuron*, *37*, pp. 159 – 170 .

Watson, J. (2002) *The phonology and morphology of Arabic*. Oxford: Oxford University Press.

White, L. (2003a) ‘Fossilization in steady state L2 grammars: Persistent problems with inflectional morphology’, *Bilingualism*: *Language and cognition,* 6(2), pp. 129-141.

White, L. (2003b) ‘Second language acquisition and Universal Grammar’, New York: Cambridge University Press.

Appendices

Appendix A: Participant Information Sheet

School of English Literature, Language and Linguistics
Percy Building
University of Newcastle upon Tyne
NE1 7RU
United Kingdom
Telephone: +44 (0)191 222 7625
Fax: +44 (0)191 222 8708

Participant Information Sheet

Acquisition of English as a Second Language

The data collected from this research project forms part of a PhD thesis, which aims to examine the acquisition of certain properties of the English language by native and non-native speakers. Previous research on the issue under investigation has mostly focused on speakers' oral production of verbal morphology neglecting perception and processing data to some extent. In this study, the three types of data will be collected. Special tests are designed to trigger data that is most appropriate to the objectives of the study. Data will be collected by the following methods:

1- You will be interviewed and the oral production will be recorded. In this interview, you will be given a questionnaire to complete and some pictures to describe. This interview will take 10 to 15 minutes at maximum.

2- You will complete a sentence retelling task, which will be audio recorded. In this task, you will listen to 50 sentences and repeat them. Before you repeat the sentences, you will be asked to choose one picture from three pictures presented on a computer screen. This task will take 20 to 25 minutes at maximum to complete.

3- You will complete a computerised picture-choice task. In this task, you will listen to 88 sentences and choose one of the pictures presented on a computer screen. An eye tracking equipment will be used in this task. It will take 35 minutes at maximum to complete this task.

Your generosity and willingness to participate in this study are greatly appreciated. Your input will help contribute to the advancement of the field of language acquisition. The data collected from you will be absolutely confidential. Your name will be converted to a code number, and only people who are associated with this research will see your name or your responses.

If you have any complaints, concerns, or questions about this research, you may contact the research supervisor, Dr. Martha Young-Scholten at martha.young-scholten@ncl.ac.uk.

Dear participant, it is your right to stop your participation in this experiment and withdraw your consent at any time you feel you need to do so.

Name of the researcher: Walid Kahoul Research supervisor: Dr. Martha Young-Scholten

Appendix B: Participant Consent Form

School of English Literature, Language and Linguistics
Percy Building
University of Newcastle upon Tyne
NE1 7RU
United Kingdom
Telephone: +44 (0)191 222 7625
Fax: +44 (0)191 222 8708

RESEARCH CONSENT FORM

Research for PhD Project: Acquisition of English as a Second Language

Researcher: Walid Kahoul

The data collected from this research project forms part of a PhD thesis, which aims to examine the acquisition of certain properties of the English language by native and non-native speakers. My involvement will consist of providing data through a questionnaire and three experimental tasks. All the data that I provide- sound files, transcripts and writing- will be anonymised, with all references to proper nouns (i.e. identifying people, places or institutions) removed. All interview data will be treated as personal under the UK 1998 Data Protection Act, and will be stored securely. The data collected as part of my participation will only be used in accordance with the permission that I give to the researcher in this form

Please tick

☐ I understand I can withdraw my consent at any time by contacting the researcher.

☐ I give my permission for the data which I will provide to be used for research purposes only (including research publications, reports, seminars).

☐ I agree to the researcher making a sound recording and to notes being taken during my interview.

☐ I confirm that I have read and understood the Information Sheet provided to me and have had the opportunity to ask the researcher questions regarding the project.

☐ I hereby assign the copyright of my contribution to the researcher.

Participant's full name:……………………………………………..

Signature:………………………………………………..

Date:…………………………………………

Appendix C: Participant Questionnaire

Participant Profile

The information you give will be treated as confidential and will only be used in data analysis. Your anonymity will be retained in the presentation of results from the study.

1. Your name:

2. Your native language(s):

3. Are you: (a) female (b) male (tick one)

4. Your date of birth:

5. Age at which you first started learning English (write native if you are a native speaker, and go to question 8)

6. Number of years you have attended English classes:

7. Have you lived in an English-speaking community? (a) yes (b) no (tick one)

If your answer is 'yes', how long in months:

8. Other languages you speak fluently:

9. Other languages you speak moderately:

10. Are you:

An undergraduate student?

A postgraduate student?

Other (please specify):

Appendix D: Proficiency Measure

D.1. Example pictures for the picture description task (Source: Levy, 1999)

D.2. Proficiency Score Results

Proficiency Data for Arab Participants

Participant	T-units	All Verbs	Different Vocabulary Items	All Words	Error-free T-units	Morphosyntactic and Lexical Accuracy	Morphosyntactic Complexity		Lexical Complexity		Proficiency Score (Factor Analysis)	Proficiency Level
							Score	Z-score	Score	Z-Score		
A1	69	109	151	432	54	0.78	1.58	-1.133	7.26	0.009	-0.491	Mid
A2	56	93	150	452	40	0.71	1.66	-0.889	7.06	-0.252	-0.737	Low
A3	63	160	221	650	49	0.78	2.54	1.798	8.67	1.847	1.941	High
A4	31	74	122	328	24	0.77	2.39	1.340	6.74	-0.669	0.293	Mid
A5	59	90	111	438	46	0.78	1.53	-1.286	5.30	-2.546	-1.982	Low
A6	66	100	119	395	59	0.89	1.52	-1.317	5.99	-1.646	-1.153	Low
A7	87	176	189	708	84	0.97	2.02	0.210	7.10	-0.199	0.636	High
A8	47	84	125	350	32	0.68	1.79	-.492	6.68	-0.747	-0.915	Low
A9	51	96	134	351	35	0.69	1.88	-.217	7.15	-0.134	-0.411	Mid
A10	56	128	165	551	50	0.89	2.29	1.035	7.03	-0.291	0.732	High
A11	77	140	209	708	65	0.84	1.82	-.400	7.85	0.778	0.477	Mid
A12	69	177	219	766	56	0.81	2.57	1.890	7.91	0.856	1.529	High
A13	61	127	149	506	46	0.75	2.08	.393	6.62	-0.825	-0.313	Mid
A14	78	154	181	795	68	0.87	1.97	.057	6.42	-1.086	-0.243	Mid
A15	51	97	181	593	24	0.47	1.90	-.156	7.43	0.231	-0.870	Low
A16	71	118	188	579	56	0.79	1.66	-.889	7.81	0.726	0.056	Mid
A17	44	91	135	379	28	0.64	2.07	.362	6.93	-0.421	-0.448	Mid
A18	78	180	205	799	51	0.65	2.31	1.096	7.25	-0.004	0.168	Mid
A19	60	144	215	796	34	0.57	2.40	1.371	7.62	0.478	0.317	Mid
A20	57	128	170	518	48	0.84	2.25	0.912	7.47	0.283	0.834	High
A21	70	128	175	560	41	0.59	1.83	-0.370	7.40	0.192	-0.618	Low
A22	59	90	152	421	45	0.76	1.53	-1.286	7.41	0.205	-0.518	Low
A23	59	115	168	553	50	0.85	1.95	-0.003	7.14	-0.147	0.186	Mid
A24	77	124	172	614	63	0.82	1.61	-1.042	6.94	-0.408	-0.553	Low
A25	58	116	180	534	41	0.71	2.00	.149	7.79	0.700	0.291	Mid
A26	63	146	193	642	40	0.63	2.32	1.126	7.62	0.478	0.387	Mid
A27	57	71	120	389	39	0.68	1.25	-2.141	6.08	-1.529	-2.143	Low
A28	59	99	165	495	40	0.68	1.68	-.828	7.42	0.218	-0.542	Low
A29	75	149	210	521	75	1.00	1.99	0.118	9.20	2.538	2.206	High
A30	52	111	175	422	44	0.85	2.13	0.546	8.52	1.652	1.449	High
A31	56	114	192	596	49	0.88	2.04	.271	7.86	0.791	0.933	High
A32	68	104	141	504	40	0.59	1.53	-1.286	6.28	-1.268	-1.869	Low
A33	74	148	181	580	64	0.86	2.00	.149	7.52	0.348	0.566	High
A34	48	108	155	472	33	0.69	2.25	.912	7.13	-0.160	0.118	Mid

Proficiency Data for Chinese Participants

Participant	T-units	All Verbs	Different Vocabulary Items	All Words	Error-free T-units	Morphosyntactic and Lexical Accuracy	Morphosyntactic Complexity		Lexical Complexity		Proficiency Score (Factor Analysis)	Proficiency Level
							Score	Z-score	Score	Z-Score		
CH1	41	79	128	373	28	0.68	1.93	0.065	6.63	-1.688	-1.169	Low
CH2	42	89	137	412	30	0.71	2.12	0.735	6.75	-1.500	-0.648	Low
CH3	34	82	160	373	21	0.62	2.41	1.758	8.28	0.901	0.894	High
CH4	44	88	141	405	31	0.70	2.00	0.312	7.01	-1.092	-0.657	Low
CH5	51	77	141	339	35	0.69	1.51	-1.416	7.66	-0.072	-0.954	Low
CH6	65	125	217	582	55	0.85	1.92	0.030	8.99	2.015	1.402	High
CH7	41	81	140	361	29	0.71	1.98	0.241	7.37	-0.527	-0.346	Mid
CH8	67	101	157	507	51	0.76	1.51	-1.416	6.97	-1.155	-1.335	Low
CH9	91	190	227	772	82	0.90	2.09	0.629	8.17	0.728	1.133	High
CH10	72	125	183	591	56	0.78	1.74	-0.605	7.53	-0.276	-0.395	Mid
CH11	75	171	206	663	63	0.84	2.28	1.299	8.00	0.461	1.119	High
CH12	56	101	165	437	46	0.82	1.80	-0.394	7.89	0.289	0.146	Mid
CH13	72	128	188	537	60	0.83	1.78	-0.464	8.11	0.634	0.335	Mid
CH14	61	119	186	596	45	0.74	1.95	0.135	7.62	-0.135	-0.085	Mid
CH15	62	117	176	521	43	0.69	1.89	-0.076	7.71	0.006	-0.266	Mid
CH16	59	103	146	366	49	0.83	1.75	-0.570	7.63	-0.119	-0.134	Mid
CH17	58	123	199	475	47	0.81	2.12	0.735	9.13	2.234	1.738	High
CH18	54	136	211	610	39	0.72	2.52	2.146	8.54	1.309	1.621	High
CH19	58	98	192	543	47	0.81	1.69	-0.782	8.24	0.838	0.233	Mid
CH20	71	141	209	632	66	0.93	1.99	0.276	8.31	0.948	1.179	High
CH21	67	132	200	570	57	0.85	1.97	0.206	8.38	1.058	0.956	High
CH22	69	112	165	531	60	0.87	1.62	-1.028	7.16	-0.857	-0.638	Low
CH23	58	97	135	377	46	0.79	1.67	-0.852	6.95	-1.186	-0.987	Low
CH24	50	95	156	449	38	0.76	1.90	-0.041	7.36	-0.543	-0.334	Mid
CH25	50	117	165	557	41	0.82	2.34	1.511	6.99	-1.123	0.279	Mid
CH26	99	173	218	773	88	0.89	1.75	-0.570	7.84	0.210	0.237	Mid
CH27	61	160	253	799	50	0.82	2.62	2.498	8.95	1.952	2.461	High
CH28	65	113	170	421	61	0.94	1.74	-0.605	8.29	0.916	0.769	High
CH29	58	117	179	554	41	0.71	2.02	0.382	7.60	-0.166	-0.078	Mid
CH30	36	75	147	404	21	0.58	2.08	0.594	7.31	-0.621	-0.636	Low
CH31	68	152	190	640	64	0.94	2.24	1.158	7.51	-0.307	0.938	High
CH32	65	104	190	576	44	0.68	1.60	-1.099	7.92	0.336	-0.606	Low
CH33	70	115	186	598	53	0.76	1.64	-0.958	7.61	-0.151	-0.557	Low
CH34	70	122	175	564	55	0.79	1.74	-0.605	7.37	-0.527	-0.503	Mid
CH35	73	107	170	535	50	0.68	1.47	-1.557	7.35	-0.559	-1.323	Low
CH36	77	130	161	572	56	0.73	1.69	-0.782	6.73	-1.531	-1.333	Low
CH37	50	83	172	561	23	0.46	1.66	-0.887	7.26	-0.700	-1.769	Low

Appendix E: The Production Study (The Sentence Elicited Imitation Task)

E.1. Example picture-sentence trial from the Sentence Elicited Imitation Task

(Picture source: http://www.google.com/imghp)

Example 1: (oral sentence) Anna goes to the zoo every week-end and stays in there all day

Example 2. (oral sentence) Yesterday, Bob played in a football match and broke his arm

E.2. Verbal Stimuli used in the Sentence Elicited Imitation Task

Practice items = 5 sentences; experimental items = 45 sentences (three batteries with 15 sentences each) (inflected and uninflected verbs are n boldface)

Practice items

1- Jack and Anna **went** to the library yesterday

2- Every day, Jack **plays** football with his friends

3- Yesterday, Jack and Anna **watched** a movie in the cinema.

4- Every year, Jack **travels** to one country and **learn** its language.

5- Two days ago, the police **caught** two criminals and **take** them to prison.

Experimental items

Battery 1:

1- Every morning, John **gets** up very early.

2- Last year, John **travelled** around several countries

3- Look! John **seem** happy today unlike yesterday.

4- Sam usually **reads** two stories before falling asleep.

5- Yesterday, Sam **borrowed** some money and **bought** a computer.

6- Sam **play** chess every day but he always **loses**.

7- Tom always **washes** the dishes and **cleans** up the kitchen.

8- Yesterday, Tom **walked** alone to school and **arrived** on time.

9- Last night, the police **catch** two thieves robbing a bank.

10- Every year, Bob **buys** a new car and **sells** the old one.

11- Last week, Bob **had** a party and **invited** all of his friends.

12- Yesterday, Tom **ask** a friend for money but he did not give him any.

13- Anna **goes** to the zoo every week-end and **stays** in there all day.

14-Sam **cried** very much yesterday because he **needed** money and nobody **gave** him.

15- Last year, Jack **tried** very hard to learn to play the guitar but he couldn't.

Experimental items

Battery 2:

1- Nick **learns** English and French at school.

2- Tom and Bob **visited** a friend yesterday.

3- Last Friday, Sam **stay** at home all day.

4- He **meets** Tom and Bob every day at school.

5- Last month, Tom **spent** only eight days in London

6- The police **stopped** me and **take** me to prison

7- Tom **misses** Jane very much and **wants** to see her

8- Tom **called** his mother yesterday and **asked** about his father.

9- Tom **saw** his teacher in the street yesterday and **talk** with her

10- Sam usually **watches** TV for twenty minutes before going to work

11- Last term, Ben **study** very hard and **passed** all the exams.

12- Tom **feel** very sad these days because of his father's death.

13- Sam has two dogs and two cats and he **feeds** them every morning

14- Tom **lived** in Mexico for ten years and another ten years in Paris

15- Every night, Tom **goes** to bed very early but he always **wake** up very late

Experimental items

Battery 3:

1- Every day, Sarah **finishes** work very early.

2- John **wanted** some help from his teacher.

3- Last year, Tom **kill** a mouse with poison.

4- John **likes** roses but **hates** butterflies.

5-Yesterday, John **ate** his breakfast before he **went** out.

6- Sam **stay** one week in Paris and **returned** yesterday.

7- John **owns** two big houses and two small flats.

8- Yesterday, Bob **played** in a football match and **broke** his arm.

9- Your dress **look** fantastic Emma but the shoes do not fit.

10- John **studies** physics at Oxford and **speaks** eight languages fluently.

11-Last week, Tom **painted** Emma's flat and **helped** her buy new furniture.

12- Yesterday, Bob **open** all the windows in the house to air it out.

13- Every Friday, Sam **pays** ten pounds to watch two movies in the cinema.

14- Last week, Sam **enjoyed** watching the movie in the cinema and he was happy.

15- Every day, Tom **brushes** his teeth and **take** a shower before going to school.

E.3. Raw Results from the Sentence Elicited Imitation Task

Arab participant's past tense production scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-d	CV-d	C-d	C-t	Irregular	Total
A2	Low	2/6	2/2	2/3	3/5	8/10	17/26
A5	Low	2/6	0/3	3/6	1/6	3/9	9/30
A6	Low	6/7	3/5	5/5	1/4	6/10	21/31
A8	Low	7/8	3/4	6/6	3/5	7/12	26/35
A15	Low	2/4	0/4	2/5	2/5	7/11	13/29
A21	Low	2/4	1/3	1/5	0/2	3/8	7/22
A22	Low	3/7	1/4	5/5	2/3	8/12	19/31
A24	Low	8/8	5/5	5/7	1/2	8/10	27/32
A27	Low	0/3	0/2	0/4	0/4	6/15	6/28
A28	Low	5/7	3/4	6/7	3/5	9/10	26/33
A32	Low	2/5	1/2	2/3	0/2	1/8	6/20
A1	Mid	5/7	1/3	3/6	1/6	5/11	15/33
A4	Mid	8/8	4/4	5/6	0/4	9/10	26/32
A9	Mid	7/8	5/5	6/7	2/2	10/11	30/33
A11	Mid	7/8	4/5	5/5	3/4	8/10	27/32
A13	Mid	6/8	3/3	4/6	3/4	6/8	22/29
A14	Mid	6/6	4/5	6/7	2/6	9/11	27/35
A16	Mid	3/5	2/4	3/4	1/4	8/12	17/29
A17	Mid	0/4	1/3	1/2	0/1	5/13	7/23
A18	Mid	3/6	3/5	2/5	2/6	8/11	18/33
A19	Mid	3/7	3/5	2/7	1/4	7/11	16/34
A23	Mid	10/10	4/4	6/6	5/5	6/7	31/32
A25	Mid	3/4	2/4	5/6	2/3	3/7	15/24
A26	Mid	5/7	0/3	2/5	0/2	4/6	11/23
A34	Mid	3/3	2/4	3/6	1/1	5/10	14/24
A3	High	6/6	3/4	7/7	2/2	11/12	29/31
A7	High	9/9	5/5	5/6	3/4	9/9	31/33
A10	High	8/8	5/5	7/7	5/5	9/9	34/34
A12	High	6/7	5/5	6/6	1/5	9/9	27/32
A20	High	7/7	4/5	5/5	3/5	10/10	29/32
A29	High	8/8	5/5	8/8	3/4	8/9	32/34
A30	High	8/8	4/5	6/7	5/5	8/10	31/35
A31	High	8/8	5/5	7/8	2/3	8/8	30/32
A33	High	7/8	5/5	6/7	4/4	10/10	32/34

Chinese participants' past tense production scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-d	CV-d	C-d	C-t	Irregular	Total
Ch1	Low	2/8	4/6	2/5	2/4	5/7	15/30
Ch2	Low	2/6	2/4	2/5	1/2	7/13	14/30
Ch4	Low	3/5	1/3	2/7	2/6	3/7	11/28
Ch5	Low	0/3	2/4	1/3	2/6	6/14	11/30
Ch8	Low	2/6	0/4	3/7	1/5	5/9	11/31
Ch22	Low	7/7	3/4	5/5	4/6	7/13	26/35
Ch23	Low	3/7	2/5	2/5	2/4	7/9	16/30
Ch30	Low	1/5	0/5	1/3	0/3	5/10	7/26
Ch32	Low	1/7	2/4	4/7	4/7	6/8	17/33
Ch33	Low	6/8	1/4	2/5	3/6	8/10	20/33
Ch35	Low	5/8	3/4	4/8	2/6	5/10	19/36
Ch36	Low	5/7	2/5	2/4	4/7	4/13	17/36
Ch37	Low	2/5	0/4	0/5	0/6	7/13	9/33
Ch7	Mid	6/6	2/3	4/5	4/5	6/9	22/28
Ch10	Mid	0/4	2/4	1/4	0/4	3/10	6/26
Ch12	Mid	5/7	1/4	0/3	2/5	7/11	15/30
Ch13	Mid	8/10	3/5	5/7	3/3	8/10	27/35
Ch14	Mid	2/6	3/4	5/6	2/5	7/14	19/35
Ch15	Mid	1/6	2/5	3/6	1/3	6/11	13/31
Ch16	Mid	2/6	0/2	0/6	0/2	4/8	6/24
Ch19	Mid	3/6	3/5	1/4	2/4	11/12	20/31
Ch24	Mid	4/6	0/4	5/7	4/7	6/9	19/33
Ch25	Mid	4/7	3/4	3/5	3/5	7/10	20/31
Ch26	Mid	2/7	1/4	1/7	1/5	7/11	12/34
Ch29	Mid	0/5	0/3	0/7	0/5	4/10	4/30
Ch34	Mid	3/5	3/3	3/5	3/6	7/12	19/31
Ch3	High	5/8	4/5	4/8	5/6	8/10	26/37
Ch6	High	3/8	4/5	4/5	5/5	9/12	25/35
Ch9	High	7/9	3/4	5/8	2/5	9/10	26/36
Ch11	High	4/8	4/4	4/6	4/4	7/8	23/30
Ch17	High	4/8	4/5	5/8	4/4	11/11	28/36
Ch18	High	4/7	1/2	3/7	1/6	8/11	17/33
Ch20	High	4/7	2/3	2/4	2/4	10/13	20/31
Ch21	High	5/6	2/5	3/4	3/5	7/11	20/31
Ch27	High	8/9	5/5	7/7	4/5	8/10	32/36
Ch28	High	8/8	4/4	6/6	5/5	9/11	32/34
Ch31	High	5/5	2/3	4/4	3/4	6/7	20/23

Arab participants' verbal agreement production scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-z	CV-z	C-z	C-s	Total
A2	Low	5/8	4/7	6/8	5/8	20/31
A5	Low	1/8	2/6	2/7	0/9	5/30
A6	Low	4/7	5/6	5/7	8/11	22/31
A8	Low	5/7	3/5	7/7	6/7	21/26
A15	Low	4/6	1/6	1/6	4/8	10/26
A21	Low	0/4	1/3	1/8	1/9	3/24
A22	Low	7/7	4/5	6/9	7/9	24/30
A24	Low	1/7	3/6	3/7	1/9	8/29
A27	Low	0/5	3/5	2/4	0/7	5/21
A28	Low	4/8	5/5	5/8	6/9	20/30
A32	Low	0/5	2/4	0/3	0/8	2/20
A1	Mid	2/4	2/5	1/6	3/7	8/22
A4	Mid	5/6	3/5	4/6	7/9	19/26
A9	Mid	6/8	5/6	1/3	6/7	18/24
A11	Mid	6/6	3/5	2/6	3/9	14/26
A13	Mid	6/7	5/6	6/7	5/10	22/30
A14	Mid	3/6	1/4	3/10	3/7	10/27
A16	Mid	1/5	0/3	3/9	3/6	7/23
A17	Mid	1/4	2/4	3/7	2/7	8/22
A18	Mid	1/6	1/2	1/9	0/6	3/23
A19	Mid	2/6	1/2	2/8	1/7	6/23
A23	Mid	5/6	6/6	6/8	7/8	24/28
A25	Mid	4/7	1/3	1/5	3/8	9/23
A26	Mid	1/6	1/4	4/5	5/7	11/22
A34	Mid	3/6	1/3	4/8	3/6	11/23
A3	High	3/3	4/5	6/7	4/6	17/21
A7	High	5/5	6/6	7/8	9/9	27/28
A10	High	7/7	5/7	4/6	6/9	22/29
A12	High	5/6	5/6	5/8	10/10	25/30
A20	High	8/8	4/6	7/9	7/7	26/30
A29	High	6/7	6/6	8/8	9/9	29/30
A30	High	7/8	5/6	8/8	6/9	26/31
A31	High	8/9	5/5	7/7	10/10	30/31
A33	High	9/9	4/6	5/6	6/9	24/30

Chinese participants' past tense production scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-z	CV-z	C-z	C-s	Total
Ch1	Low	2/5	4/5	3/7	9/10	18/27
Ch2	Low	5/9	1/5	4/8	4/9	14/31
Ch4	Low	5/7	4/4	3/8	5/9	17/28
Ch5	Low	2/3	1/3	2/7	2/7	7/20
Ch8	Low	2/7	2/6	4/6	7/9	15/28
Ch22	Low	3/6	5/7	6/9	8/9	22/31
Ch23	Low	2/6	3/6	4/8	6/9	15/29
Ch30	Low	1/4	3/6	1/7	2/6	7/23
Ch32	Low	5/9	3/5	2/6	5/9	15/29
Ch33	Low	2/6	2/6	4/7	8/8	16/27
Ch35	Low	4/6	3/5	5/7	5/11	17/29
Ch36	Low	3/6	3/5	4/6	5/9	15/26
Ch37	Low	4/5	2/5	4/6	4/10	14/26
Ch7	Mid	5/5	4/5	7/10	7/8	23/28
Ch10	Mid	2/6	2/5	4/8	4/8	12/27
Ch12	Mid	5/8	3/6	3/9	5/7	16/30
Ch13	Mid	4/7	1/4	5/7	7/9	17/27
Ch14	Mid	1/4	2/4	2/7	3/8	8/23
Ch15	Mid	4/8	0/6	1/8	4/11	9/33
Ch16	Mid	0/3	0/5	1/6	0/9	1/23
Ch19	Mid	6/8	4/7	5/7	7/8	22/30
Ch24	Mid	3/6	2/6	4/7	5/8	14/27
Ch25	Mid	3/5	1/6	3/6	5/11	12/28
Ch26	Mid	0/7	3/4	3/6	1/9	7/26
Ch29	Mid	06	2/5	1/5	1/5	4/21
Ch34	Mid	4/8	5/6	6/6	8/9	23/29
Ch3	High	5/7	3/5	3/7	6/7	17/26
Ch6	High	4/7	4/6	7/9	6/8	21/30
Ch9	High	4/8	6/6	5/8	8/9	23/31
Ch11	High	6/7	6/6	7/8	7/9	26/30
Ch17	High	1/3	6/6	5/7	6/7	18/23
Ch18	High	5/7	5/5	3/9	5/10	18/31
Ch20	High	3/6	3/6	4/5	3/7	13/24
Ch21	High	4/7	5/6	4/9	7/9	20/31
Ch27	High	4/5	5/6	6/8	6/7	21/26
Ch28	High	7/7	6/6	7/7	8/8	28/28
Ch31	High	1/2	3/4	4/6	4/6	12/18

Arab participants' past production by prosodic structures (supplied/obligatory contexts)

Participant No	Proficiency Level	Inflection as Onset "lived on" [levdon]	Inflection as Coda "visited" [visitəd]	Inflection as Foot-internal "talked" [tokt] or "played" [pleid].	No Option for inflection within PWd
A2	Low	1/3	2/2	6/11	0
A5	Low	0	0/3	6/16	0/3
A6	Low	3/5	3/5	9/12	0
A8	Low	1/2	3/4	13/17	0
A15	Low	3/5	0/4	3/9	0
A21	Low	0/3	1/4	3/8	0
A22	Low	4/4	1/4	6/11	0
A24	Low	2/4	5/6	9/13	0
A27	Low	0/2	0/2	0/7	0/2
A28	Low	3/3	3/4	10/16	1/1
A32	Low	0	1/2	4/9	0/1
A1	Mid	0/2	1/3	8/14	1/3
A4	Mid	1/3	4/4	11/14	1/1
A9	Mid	3/3	5/5	11/13	1/1
A11	Mid	3/4	4/5	12/13	0
A13	Mid	2/2	3/3	10/15	1/1
A14	Mid	2/7	4/5	12/12	0
A16	Mid	1/3	2/4	6/10	0
A17	Mid	0/1	1/3	1/6	0
A18	Mid	1/4	3/5	6/13	0
A19	Mid	2/4	3/5	4/13	0/1
A23	Mid	5/5	4/4	15/15	1/1
A25	Mid	3/3	2/4	6/9	1/1
A26	Mid	0/3	0/3	7/11	0
A34	Mid	0/2	2/4	7/7	0/1
A3	High	1/1	3/4	12/12	2/2
A7	High	4/5	5/5	13/14	0
A10	High	5/5	5/5	14/14	1/1
A12	High	3/5	5/5	10/13	0
A20	High	3/5	4/5	12/12	0
A29	High	4/5	5/5	15/15	0
A30	High	4/5	4/5	15/15	0
A31	High	3/4	5/5	14/15	0
A33	High	5/5	5/5	12/14	0

Chinese participants' past production by prosodic structures (supplied/obligatory contexts)

Participant No	Proficiency Level	Inflection as Onset "lived on" [levdon]	Inflection as Coda "visited" [visitəd]	Inflection as Foot-internal "talked" [tokt] or "played" [pleid].	No Option for inflection within PWd
Ch1	Low	2/3	4/6	3/13	1/1
Ch2	Low	0/1	2/4	5/12	0
Ch4	Low	1/5	1/3	5/13	0
Ch5	Low	0/2	2/4	1/6	1/4
Ch8	Low	0/3	0/4	8/15	0
Ch22	Low	1/2	3/4	12/14	2/2
Ch23	Low	0/2	2/5	6/13	0/1
Ch30	Low	0/2	0/5	2/9	0
Ch32	Low	3/3	2/4	6/15	2/3
Ch33	Low	1/1	1/4	9/15	1/3
Ch35	Low	1/5	3/4	10/17	0
Ch36	Low	2/3	2/5	9/14	0/1
Ch37	Low	0/3	0/4	2/12	0/1
Ch7	Mid	1/2	2/3	12/13	0
Ch10	Mid	0/2	2/4	1/11	0/1
Ch12	Mid	2/4	1/4	5/11	0
Ch13	Mid	3/3	3/5	13/17	0
Ch14	Mid	1/3	3/4	7/13	1/1
Ch15	Mid	2/3	2/5	3/12	0
Ch16	Mid	0/1	0/2	2/11	0/2
Ch19	Mid	0/4	3/5	6/10	0
Ch24	Mid	3/3	0/4	9/14	1/3
Ch25	Mid	1/3	3/4	9/13	0/1
Ch26	Mid	1/5	1/4	2/13	1/1
Ch29	Mid	0/2	0/5	0/11	0/2
Ch34	Mid	2/4	3/3	7/11	0/1
Ch3	High	2/4	4/5	11/16	1/2
Ch6	High	2/4	4/5	10/16	0
Ch9	High	1/4	3/4	13/17	0/1
Ch11	High	2/4	4/4	10/14	0
Ch17	High	4/5	4/5	9/15	0
Ch18	High	1/4	1/2	6/14	1/2
Ch20	High	1/3	2/3	7/10	0/2
Ch21	High	2/2	2/5	9/11	0/2
Ch27	High	4/4	5/5	15/16	01
Ch28	High	4/4	4/4	14/14	1/1
Ch31	High	1/1	2/3	10/10	1/2

Arab participants' verbal agreement production by prosodic structures (supplied/obligatory contexts)

Participant No	Proficiency Level	Inflection as Onset "builds on" [bIldzan]	Inflection as Coda "races" [rejsəz]	Inflection as Foot-internal "fills" [fIlz] or "sews" [sowz].	No Option for Agreement within PWd
A2	Low	0	4/7	11/17	5/7
A5	Low	0	2/6	2/16	1/8
A6	Low	0	5/6	14/17	3/8
A8	Low	0	3/5	12/14	6/7
A15	Low	0	1/6	8/15	1/5
A21	Low	0	1/3	0/12	2/9
A22	Low	0	4/5	13/14	7/11
A24	Low	0	3/6	2/16	3/7
A27	Low	0	3/5	0/10	2/6
A28	Low	0	5/5	9/17	6/8
A32	Low	0	2/4	0/14	0/2
A1	Mid	0	2/5	5/12	1/5
A4	Mid	0	3/5	12/16	4/5
A9	Mid	0	5/6	11/14	2/4
A11	Mid	0	3/5	9/14	2/7
A13	Mid	0	5/6	12/17	5/7
A14	Mid	0	1/4	5/13	4/10
A16	Mid	0	0/3	3/10	4/10
A17	Mid	0	2/4	5/11	1/7
A18	Mid	0	1/2	2/14	0/7
A19	Mid	0	1/2	2/13	3/8
A23	Mid	0	6/6	13/15	5/7
A25	Mid	0	1/3	6/15	2/5
A26	Mid	0	1/4	7/13	3/5
A34	Mid	0	1/3	9/15	1/5
A3	High	0	4/5	9/11	4/5
A7	High	0	6/6	14/14	7/8
A10	High	0	5/7	10/14	7/8
A12	High	0	5/6	15/16	5/8
A20	High	0	4/6	15/15	7/9
A29	High	0	6/6	15/16	8/8
A30	High	0	5/6	15/19	6/6
A31	High	0	5/5	16/17	9/9
A33	High	0	4/6	15/18	5/6

Chinese participants' verbal agreement production by prosodic structures (supplied/obligatory contexts)

Participant No	Proficiency Level	Inflection as Onset "builds on" [bIldzan]	Inflection as Coda "races" [rejsəz]	Inflection as Foot-internal "fills" [fIlz] or "sews" [sowz].	No Option for Agreement within PWd
Ch1	Low	0	4/5	12/15	2/7
Ch2	Low	0	1/5	11/23	4/7
Ch4	Low	0	4/4	7/11	4/9
Ch5	Low	0	1/3	4/12	2/5
Ch8	Low	0	2/6	11/18	2/4
Ch22	Low	0	5/7	10/14	7/10
Ch23	Low	0	3/6	9/15	3/8
Ch30	Low	0	3/6	2/10	2/7
Ch32	Low	0	3/5	10/17	2/7
Ch33	Low	0	2/6	9/13	5/8
Ch35	Low	0	3/5	10/16	4/8
Ch36	Low	0	3/5	8/15	4/6
Ch37	Low	0	2/5	10/15	2/6
Ch7	Mid	0	4/5	12/14	7/9
Ch10	Mid	0	2/5	9/17	1/5
Ch12	Mid	0	3/6	11/18	2/6
Ch13	Mid	0	1/4	12/16	4/7
Ch14	Mid	0	2/4	4/13	2/6
Ch15	Mid	0/1	0/6	8/19	1/7
Ch16	Mid	0	0/5	1/11	1/7
Ch19	Mid	0	4/7	13/17	4/6
Ch24	Mid	0	2/6	10/16	2/5
Ch25	Mid	0	1/6	8/14	3/8
Ch26	Mid	0	3/4	3/16	1/6
Ch29	Mid	0	2/5	2/13	0/3
Ch34	Mid	0	5/6	11/16	7/7
Ch3	High	0	3/5	10/12	4/9
Ch6	High	0	4/6	10/16	7/8
Ch9	High	0	6/6	12/17	5/8
Ch11	High	0	6/6	13/16	7/8
Ch17	High	0	6/6	8/11	4/6
Ch18	High	0	5/5	11/17	2/9
Ch20	High	0	3/6	5/12	5/6
Ch21	High	0/1	5/6	10/16	5/8
Ch27	High	0	5/6	10/14	6/6
Ch28	High	0	6/6	15/15	7/7
Ch31	High	0	3/4	4/7	5/7

Appendix F: Perception and Processing Study (The Computerised Picture Choice Task)

F.1. Example picture-sentence trial from the Computerised Picture Choice Task

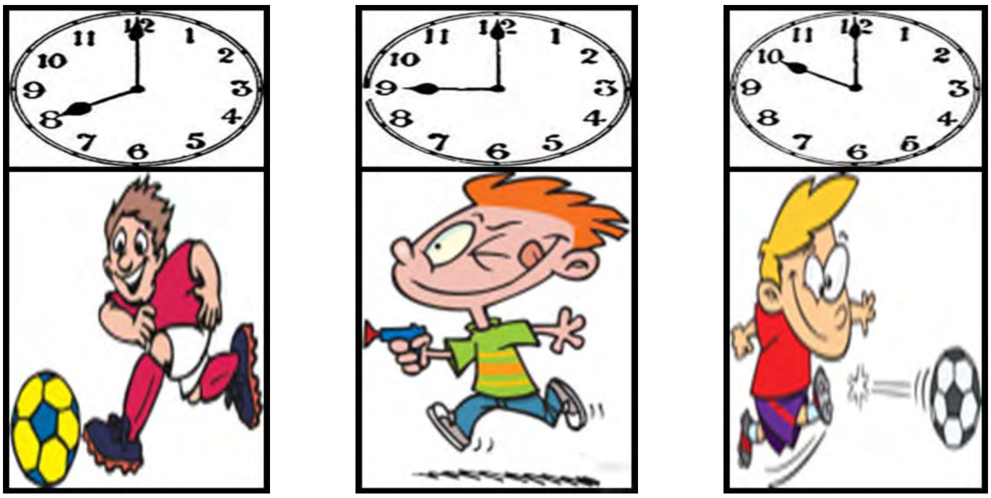
(Picture source: http://www.google.com/imghp)

Example 1: Pair of trials (distractor/ experimental) testing past tense

a: Distractor Trial

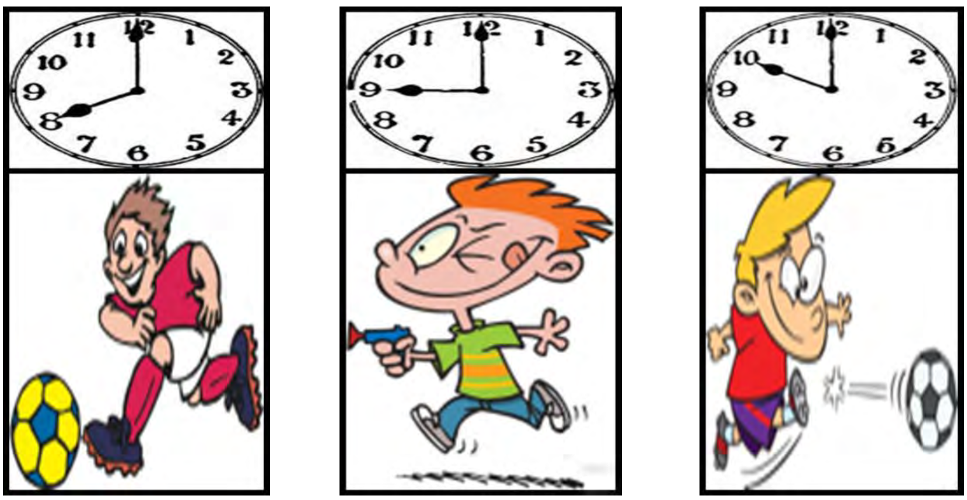
(Oral presentation) It is 10 o'clock right now. Listen and choose:

"He was playing one hour ago."



b: Experimental Trial

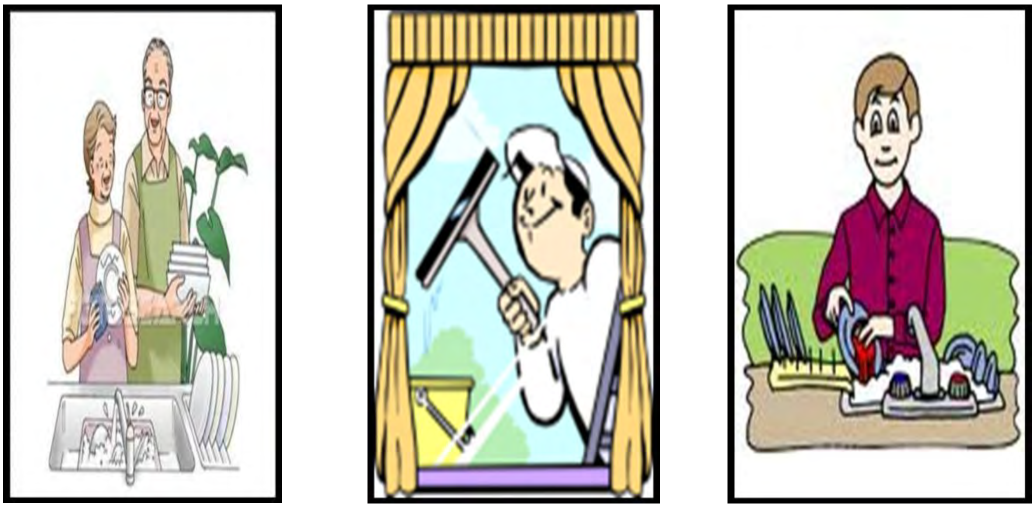
(Oral presentation) "He broke his arm while playing football



Example 2: Pair of trials (distractor/ experimental) testing verbal agreement

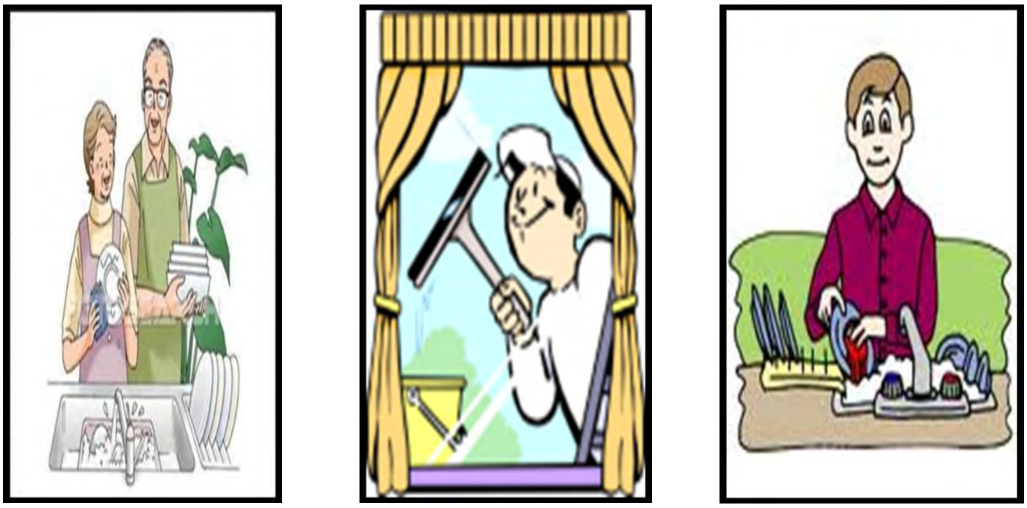
a: Distractor Trial

(Oral presentation) “He is cleaning the window”



b: Experimental Trial

(Oral presentation) “Everyday......washes up the dishes**.”**



F.2: The verbal stimuli used in the computerised picture choice task.

(Practice trials = 10; test trials = 78 (total = 88))

Practice Trials:

1- It is 10 o'clock right now. Listen to the following and choose only one picture: What time is it now?

2- Listen and choose one picture: They are very nice.

3- It is 10 o'clock right now. Listen to the following and choose only one picture: What time was it one hour ago?

4- She is beautiful.

5- It is 10 o'clock right now. Listen and choose one picture: She was playing tennis two hours ago

6- They are playing football.

7- It is 10 o'clock right now. Listen and choose one picture: She is playing tennis now.

8- He is playing football.

9- They seem good friends in the picture.

10- It is 10 o'clock right now. Listen and choose: He was eating sometime ago.

Test Trials: (Researcher version, i.e., nonrandomised items)

11- Everyday …watches TV.

12- She likes video games.

13- It is 10 o'clock right now. Listen and choose: He was playing one hour ago.

14- He broke his arm while playing football.

15- They are having fun.

16- Everyday...reads for two hours.

17- It is 10 o'clock right now. Listen and choose: She was doing something two hours ago.

18- She cleaned up the whole place.

19- He is very happy.

20- While running…gets tired quickly.

21- It is 10 o'clock right now. Listen and choose: This person was typing one hour ago.

22- She typed eight pages before going home.

23- Every morning…goes to school very early.

24- The musician is having fun.

25- It is 10 o'clock right now. Listen and choose: It happened two hours ago.

26- The baby kept crying until the mother's arrival.

27- It is 10 o'clock right now. Listen and Choose: They were together two hours ago.

28- They played a chess game together.

29- They clean the place every day.

30- Everyday...swims for half an hour.

31- It is 10 o'clock right now. Listen and choose: It happened one hour ago.

32- They stayed playing on the bed for half an hour.

33- Every party...dances happily.

34- They are jumping all together.

35- It is 10 o'clock right now. Listen and choose: They were together one hour ago

36- They waited for a long time before getting on the bus.

37- Every party…sings happily**.**

38- He is playing the guitar.

39- It is 10 o'clock right now. Listen and choose: He was climbing something one hour ago.

40- He climbed up the mountain.

41- Every winter...skies in the snow.

42- Look! They are playing together.

43- It is 10 o'clock right now. Listen and choose: That happened two hours ago.

44- She cried very hard."

45- They are playing basketball.

46- Every day…plays music for two hours.

47- It is 10 o'clock right now. Listen and choose: That was one hour ago.

48- They went to school together.

49- Every weekend...drinks a lot of wine.

50- They are not drinking wine.

51- It is 10 o'clock right now. Listen and choose: That happened one hour ago.

52- They watched a funny show on TV.

53- He is cleaning the window.

54- Everyday...washes up the dishes.

55- It is 10 o'clock right now. Listen and choose: It happened two hours ago.

56- He phoned a friend before going out.

57- It is 10 o'clock right now. Listen and choose: They were together one hour ago.

58- They walked over the bridge.

59- It is 10 o'clock right now. Listen and choose: He was there two hours ago.

60- He painted a wall.

61- He is cutting meat.

62- Once a week...eats at a restaurant.

63- It is 10 o'clock right now. Listen and choose: "He was carrying something two hours ago.

64- He lifted up the heavy weights.

65- Every weekend...drinks a lot of wine.

66- They are not drinking wine.

67- He is cleaning the window.

68- Everyday...wash up the dishes

69- He is cutting meat.

70- Once a week...eat at a restaurant.

71- Everyday …watch TV.

72- She likes video games.

73-They are having fun.

74- Everyday...read for two hours

75- Every morning…go to school very early.

76- The musician is having fun.

77- It is 10 o'clock right now. Listen and choose: That happened one hour ago.

78- They watch a funny show on TV.

79- It is 10 o'clock right now. Listen and choose: They were together one hour ago.

80- They walk over the bridge.

81- It is 10 o'clock right now. Listen and choose: He was there two hours ago.

82- He paint a wall

83- It is 10 o'clock right now. Listen and choose: She was doing something two hours ago.

84- She clean up the whole place.

85- It is 10 o'clock right now. Listen and Choose: They were together two hours ago.

86- They play a chess game together.

87- It is 10 o'clock right now. Listen and choose: It happened one hour ago.

88- They stay playing on the bed for half an hour.

F.3. Raw Results from the Computerised Picture Choice Task: Behavioural Response

Native speakers' past tense perception scores by allomorphs (supplied/obligatory contexts)

Participant NO	VV-d	CV-d	C-d	C-t	Irregular	Total
NS1	3/3	3/3	3/3	3/3	3/3	15/15
NS2	3/3	2/3	3/3	2/3	2/3	12/15
NS3	3/3	3/3	3/3	3/3	3/3	15/15
NS4	3/3	3/3	2/3	2/3	2/3	12/15
NS5	3/3	3/3	3/3	3/3	3/3	15/15
NS6	3/3	3/3	3/3	3/3	3/3	15/15
NS7	2/3	3/3	3/3	3/3	2/3	13/15
NS8	3/3	3/3	3/3	3/3	3/3	15/15
NS9	3/3	3/3	3/3	3/3	3/3	15/15
NS10	2/3	2/3	3/3	2/3	3/3	12/15

Arab participant's past tense perception scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-d	CV-d	C-d	C-t	Irregular	Total
A2	Low	1/3	1/3	2/3	2/3	2/3	8/15
A5	Low	1/3	1/3	1/3	3/3	1/3	7/15
A6	Low	1/3	3/3	3/3	3/3	2/3	12/15
A8	Low	2/3	1/3	1/3	1/3	1/3	6/15
A15	Low	.	.	.	.	.	.
A21	Low	.	.	.	.	.	.
A22	Low	3/3	2/3	2/3	2/3	2/3	12/15
A24	Low	3/3	2/3	3/3	2/3	2/3	12/15
A27	Low	0/3	0/3	2/3	1/3	2/3	5/15
A28	Low	2/3	2/3	1/3	2/2	2/2	9/15
A32	Low	1/3	1/3	2/3	1/3	1/3	6/15
A1	Mid	2/3	3/3	2/3	2/3	0/3	9/15
A4	Mid	.	.	.	.	.	.
A9	Mid	3/3	2/3	3/3	3/3	3/3	14/15
A11	Mid	3/3	3/3	3/3	3/3	3/3	15/15
A13	Mid	3/3	2/3	3/3	3/3	3/3	14/15
A14	Mid	2/3	3/3	2/3	2/3	2/3	11/15
A16	Mid	2/3	2/3	1/3	2/3	1/3	8/15
A17	Mid	0/3	0/3	0/3	1/3	0/3	1/15
A18	Mid	2/3	3/3	1/3	2/3	3/3	11/15
A19	Mid	2/3	3/3	3/3	2/3	2/3	12/15
A23	Mid	2/3	0/3	1/3	1/3	2/3	6/15
A25	Mid	2/3	2/3	3/3	3/3	1/3	11/15
A26	Mid	3/3	3/3	3/3	2/3	2/3	13/15
A34	Mid	1/3	2/3	3/3	1/3	1/3	8/15
A3	High	2/3	3/3	3/3	3/3	3/3	14/15
A7	High	3/3	3/3	3/3	3/3	3/3	15/15
A10	High	3/3	3/3	2/3	3/3	1/3	12/15
A12	High	3/3	3/3	3/3	2/3	2/3	13/15
A20	High	2/3	3/3	2/3	2/3	3/3	12/15
A29	High	3/3	3/3	3/3	2/3	2/3	13/15
A30	High	1/3	1/3	3/3	2/3	3/3	10/15
A31	High	3/3	3/3	3/3	3/3	3/3	15/15
A33	High	3/3	3/3	3/3	2/3	2/3	13/15

Chinese participant's past tense perception scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-d	CV-d	C-d	C-t	Irregular	Total
Ch1	Low	0/3	2/3	1/3	1/3	2/3	6/15
Ch2	Low	2/3	1/1	2/2	2/2	1/3	8/11
Ch4	Low	1/3	3/3	3/3	2/3	2/3	11/15
Ch5	Low	0/3	2/3	1/3	1/3	2/3	6/15
Ch8	Low	3/3	3/3	1/3	3/3	3/3	13/15
Ch22	Low	2/3	3/3	3/3	3/3	1/3	12/15
Ch23	Low	0/3	3/3	2/3	2/3	2/3	9/15
Ch30	Low	0/3	0/3	2/3	1/3	2/3	5/15
Ch32	Low	3/3	3/3	3/3	2/3	2/3	13/15
Ch33	Low	3/3	3/3	2/3	1/3	2/3	11/15
Ch35	Low	1/3	0/3	2/3	2/3	3/3	8/15
Ch36	Low	.	.	.	.	.	.
Ch37	Low	.	.	.	.	.	.
Ch7	Mid	3/3	3/3	2/3	2/3	3/3	13/15
Ch10	Mid	2/3	1/3	2/3	2/3	2/3	9/15
Ch12	Mid	2/3	2/3	2/3	2/3	2/3	10/15
Ch13	Mid	2/3	2/3	2/3	3/3	2/3	11/15
Ch14	Mid	.	.	.	.	.	.
Ch15	Mid	1/3	1/3	1/3	1/3	0/3	4/15
Ch16	Mid	1/3	1/3	3/3	3/3	2/3	10/15
Ch19	Mid	3/3	3/3	3/3	3/3	3/3	15/15
Ch24	Mid	0/3	0/3	1/3	2/3	2/3	5/15
Ch25	Mid	1/3	1/3	1/3	2/3	2/3	7/15
Ch26	Mid	2/3	2/3	1/3	3/3	3/3	11/15
Ch29	Mid	1/3	2/3	1/3	2/3	3/3	9/15
Ch34	Mid	0/3	0/3	1/3	1/3	1/3	3/15
Ch3	High	3/3	3/3	3/3	3/3	3/3	15/15
Ch6	High	1/3	0/3	3/3	2/3	3/3	9/15
Ch9	High	3/3	3/3	1/3	2/3	2/3	11/15
Ch11	High	3/3	3/3	3/3	3/3	2/3	14/15
Ch17	High	1/3	2/3	3/3	2/3	2/3	10/15
Ch18	High	1/3	2/3	0/3	2/3	2/3	7/15
Ch20	High	1/3	1/3	1/3	0/3	3/3	6/15
Ch21	High	2/3	3/3	3/3	3/3	3/3	14/15
Ch27	High	0/3	0/3	2/3	1/3	1/3	4/15
Ch28	High	3/3	3/3	3/3	3/3	2/3	14/15
Ch31	High	.	.	.	.	.	.

Native speakers' verbal agreement perception scores by allomorphs (supplied/obligatory contexts)

Participant NO	VV-z	CV-z	C-z	C-s	Total
NS1	3/3	3/3	3/3	3/3	12/12
NS2	3/3	3/3	3/3	3/3	12/12
NS3	3/3	3/3	3/3	3/3	12/12
NS4	3/3	3/3	3/3	3/3	12/12
NS5	3/3	3/3	3/3	3/3	12/12
NS6	2/3	3/3	3/3	3/3	11/12
NS7	3/3	3/3	2/3	3/3	11/12
NS8	3/3	3/3	3/3	3/3	12/12
NS9	3/3	3/3	3/3	3/3	12/12
NS10	3/3	2/3	3/3	3/3	11/12

Arab participant's verbal agreement perception scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	VV-z	CV-z	C-z	C-s	Total
A2	Low	2/3	1/3	3/3	2/3	8/12
A5	Low	2/3	2/3	1/3	2/3	7/12
A6	Low	3/3	3/3	3/3	3/3	12/12
A8	Low	3/3	3/3	3/3	2/3	11/12
A15	Low	.	.	.	.	.
A21	Low	.	.	.	.	.
A22	Low	2/3	2/3	2/3	1/3	7/12
A24	Low	3/2	2/2	2/2	3/2	10/12
A27	Low	2/3	2/3	3/3	3/3	10/12
A28	Low	1/3	2/3	3/3	2/3	8/12
A32	Low	1/3	2/3	3/3	1/3	7/12
A1	Mid	3/3	2/3	2/3	3/3	10/12
A4	Mid	.	.	.	.	.
A9	Mid	3/3	3/3	3/3	3/3	12/12
A11	Mid	3/3	3/3	3/3	3/3	12/12
A13	Mid	1/3	1/3	2/3	2/3	6/12
A14	Mid	3/3	2/3	2/3	1/3	8/12
A16	Mid	2/3	2/3	2/3	1/3	7/12
A17	Mid	1/3	2/3	1/3	2/3	6/12
A18	Mid	1/3	2/3	2/3	2/3	7/12
A19	Mid	2/3	3/3	2/3	3/3	10/12
A23	Mid	2/3	1/3	2/3	2/3	8/12
A25	Mid	3/3	2/3	1/3	1/3	7/12
A26	Mid	3/3	3/3	3/3	2/3	11/12
A34	Mid	2/3	3/3	1/3	2/3	8/12
A3	High	3/3	3/3	3/3	3/3	12/12
A7	High	3/3	3/3	3/3	3/3	12/12
A10	High	3/3	3/3	3/3	3/3	12/12
A12	High	3/3	3/3	2/3	2/3	10/12
A20	High	3/3	2/3	3/3	1/3	9/12
A29	High	3/3	3/3	3/3	3/3	12/12
A30	High	3/3	2/3	3/3	3/3	11/12
A31	High	3/3	3/3	2/3	3/3	11/12
A33	High	3/3	3/3	3/3	2/3	11/12

Chinese participant's verbal agreement perception scores by allomorphs (supplied/obligatory contexts)

Participant NO	Proficiency Level	V-z	CV-z	C-z	C-s	Total
Ch1	Low	3/2	2/3	2/3	3/3	10/12
Ch2	Low	3/3	1/2	2/3	2/2	8/10
Ch4	Low	3/3	3/3	3/3	3/3	12/12
Ch5	Low	3/3	2/3	3/3	2/3	10/12
Ch8	Low	3/3	2/3	3/3	2/3	10/12
Ch22	Low	2/3	2/3	2/3	2/3	8/12
Ch23	Low	3/2	2/3	2/3	2/3	9/12
Ch30	Low	3/3	1/3	2/3	3/3	9/12
Ch32	Low	3/3	2/3	2/3	3/3	10/12
Ch33	Low	3/3	3/3	3/3	2/3	11/12
Ch35	Low	3/3	1/3	2/3	2/3	8/12
Ch36	Low	.	.	.	.	.
Ch37	Low	.	.	.	.	.
Ch7	Mid	3/2	2/3	3/3	2/2	10/12
Ch10	Mid	3/3	2/3	1/3	3/3	9/12
Ch12	Mid	3/3	2/3	3/3	3/3	11/12
Ch13	Mid	3/3	1/3	3/3	1/3	8/12
Ch14	Mid	3/3	2/3	3/3	3/3	11/12
Ch15	Mid	2/3	2/3	3/3	2/3	9/12
Ch16	Mid	3/3	2/3	2/3	3/3	10/12
Ch19	Mid	3/3	3/3	3/3	2/3	11/12
Ch24	Mid	3/3	2/3	2/3	2/3	9/12
Ch25	Mid	3/3	2/3	3/3	2/3	10/12
Ch26	Mid	1/3	2/3	2/3	1/3	6/12
Ch29	Mid	2/3	2/3	0/3	1/3	5/12
Ch34	Mid	2/3	2/3	2/3	3/3	9/12
Ch3	High	3/3	3/3	3/3	3/3	12/12
Ch6	High	3/3	3/3	3/3	3/3	12/12
Ch9	High	3/3	2/3	3/3	3/3	11/12
Ch11	High	2/3	2/3	3/2	2/3	9/12
Ch17	High	3/3	1/3	3/3	3/3	10/12
Ch18	High	1/3	2/3	2/3	0/3	5/12
Ch20	High	3/3	2/3	2/3	3/3	10/12
Ch21	High	2/3	2/3	2/3	1/3	7/12
Ch27	High	3/3	2/3	2/3	3/3	10/12
Ch28	High	3/3	1/3	2/3	2/3	8/12
Ch31	High	.	.	.	.	.

F.4. Results from the Computerised Picture Choice Task: Individual RT Means

Native speakers' individual means of RTs in response to past tense trials

Participant NO	**Means in Seconds**
NS1	1.443
NS2	0.845
NS3	0.796
NS4	1.159
NS5	1.043
NS6	0.998
NS7	1.123
NS8	0.774
NS9	0.844
NS10	1.146

Arab participant's individual means of RTs in response to past tense trials

Participant NO	Proficiency Level	Means in Seconds
A2	Low	3.874
A5	Low	2.952
A6	Low	1.839
A8	Low	1.749
A15	Low	
A21	Low	
A22	Low	1.046
A24	Low	2.228
A27	Low	2.999
A28	Low	2.498
A32	Low	4.829
A1	Mid	3.206
A4	Mid	
A9	Mid	1.094
A11	Mid	.887
A13	Mid	2.294
A14	Mid	1.826
A16	Mid	1.794
A17	Mid	.
A18	Mid	1.970
A19	Mid	1.839
A23	Mid	2.077
A25	Mid	2.919
A26	Mid	1.882
A34	Mid	4.050
A3	High	1.045
A7	High	1.152
A10	High	2.184
A12	High	1.255
A20	High	3.024
A29	High	1.214
A30	High	1.634
A31	High	1.144
A33	High	.876

Chinese participant's individual means of RTs in response to past tense trials

Participant NO	Proficiency Level	Means in Seconds
Ch1	Low	1.938
Ch2	Low	3.656
Ch4	Low	3.490
Ch5	Low	1.989
Ch8	Low	1.813
Ch22	Low	2.524
Ch23	Low	2.839
Ch30	Low	3.787
Ch32	Low	2.747
Ch33	Low	2.424
Ch35	Low	2.819
Ch36	Low	
Ch37	Low	
Ch7	Mid	1.699
Ch10	Mid	1.628
Ch12	Mid	2.334
Ch13	Mid	2.507
Ch14	Mid	
Ch15	Mid	2.342
Ch16	Mid	5.725
Ch19	Mid	1.833
Ch24	Mid	2.865
Ch25	Mid	4.408
Ch26	Mid	1.499
Ch29	Mid	1.769
Ch34	Mid	3.074
Ch3	High	4.656
Ch6	High	2.121
Ch9	High	2.006
Ch11	High	0.808
Ch17	High	1.818
Ch18	High	2.006
Ch20	High	1.562
Ch21	High	0.974
Ch27	High	1.790
Ch28	High	0.800
Ch31	High	

Native speakers' individual means of RTs in response to verbal agreement trials

Participant NO	Means in Seconds
NS1	1.032
NS2	0.744
NS3	0.756
NS4	0.905
NS5	0.792
NS6	0.923
NS7	0.749
NS8	1.085
NS9	0.675
NS10	0.910

Arab participant's individual means of RTs in response to verbal agreement trials

Participant NO	Proficiency Level	Means in Seconds
A2	Low	3.635
A5	Low	2.248
A6	Low	2.995
A8	Low	1.756
A15	Low	
A21	Low	
A22	Low	1.815
A24	Low	2.101
A27	Low	1.980
A28	Low	1.709
A32	Low	4.126
A1	Mid	3.321
A4	Mid	
A9	Mid	1.235
A11	Mid	.926
A13	Mid	3.710
A14	Mid	1.811
A16	Mid	1.226
A17	Mid	2.355
A18	Mid	2.261
A19	Mid	1.236
A23	Mid	1.379
A25	Mid	2.538
A26	Mid	2.431
A34	Mid	5.482
A3	High	1.163
A7	High	1.170
A10	High	2.638
A12	High	1.445
A20	High	1.117
A29	High	1.074
A30	High	1.622
A31	High	1.000
A33	High	1.189

Chinese participant's individual means of RTs in response to verbal agreement trials

Participant NO	Proficiency Level	Means in Seconds
Ch1	Low	2.056
Ch2	Low	2.874
Ch4	Low	5.411
Ch5	Low	2.560
Ch8	Low	1.878
Ch22	Low	1.846
Ch23	Low	2.193
Ch30	Low	1.745
Ch32	Low	2.868
Ch33	Low	2.177
Ch35	Low	1.547
Ch36	Low	
Ch37	Low	
Ch7	Mid	0.693
Ch10	Mid	1.262
Ch12	Mid	1.642
Ch13	Mid	1.079
Ch14	Mid	2.661
Ch15	Mid	2.987
Ch16	Mid	4.258
Ch19	Mid	1.120
Ch24	Mid	1.544
Ch25	Mid	2.312
Ch26	Mid	1.025
Ch29	Mid	1.222
Ch34	Mid	2.468
Ch3	High	4.142
Ch6	High	1.431
Ch9	High	1.426
Ch11	High	0.917
Ch17	High	1.910
Ch18	High	1.621
Ch20	High	1.405
Ch21	High	0.879
Ch27	High	0.961
Ch28	High	0.722
Ch31	High	

F.5. Results from the Computerised Picture Choice Task: Individual Means of Length of Eye Looks

Native speakers' individual means of length of eye looks to competitor and target pictures in past tense trials

Participant NO	Target looks in second	Competitor looks in second
NS1	2.701	1.447
NS2	2.343	1.367
NS3	2.682	1.153
NS4	2.212	1.541
NS5	2.035	1.401
NS6	2.049	1.171
NS7	2.096	1.585
NS8	2.352	1.479
NS9	2.311	1.293
NS10	1.736	1.915

Arab participant's individual means of length of eye looks to competitor and target pictures in past tense trials

Participant NO	Proficiency Level	Target looks in second	Competitor looks in second
A2	Low	2.153	1.405
A5	Low	2.870	2.257
A6	Low	3.175	2.179
A8	Low	1.940	2.784
A15	Low		
A21	Low		
A22	Low	1.821	2.097
A24	Low	2.592	2.321
A27	Low	1.564	1.899
A28	Low	0.995	3.086
A32	Low	3.877	3.803
A1	Mid	2.790	2.300
A4	Mid		
A9	Mid	2.840	1.146
A11	Mid	2.704	1.153
A13	Mid	1.955	2.856
A14	Mid	1.931	1.467
A16	Mid	2.179	1.922
A17	Mid		
A18	Mid	2.623	2.250
A19	Mid	2.101	1.692
A23	Mid	2.479	2.180
A25	Mid	3.413	1.746
A26	Mid	2.911	1.997
A34	Mid	3.335	2.067
A3	High	2.688	1.175
A7	High	3.117	1.174
A10	High	3.005	2.010
A12	High	2.645	1.425
A20	High	3.067	2.020
A29	High	2.283	1.906
A30	High	2.678	2.782
A31	High	2.492	1.325
A33	High	2.206	1.083

Chinese participant's individual means of length of eye looks to competitor and target pictures in past tense trials

Participant NO	Proficiency Level	Target looks in second	Competitor looks in second
Ch1	Low	1.770	2.151
Ch2	Low	2.994	3.378
Ch4	Low	2.838	2.547
Ch5	Low	2.157	2.140
Ch8	Low	2.040	1.878
Ch22	Low	1.749	2.608
Ch23	Low	2.484	2.434
Ch30	Low	2.324	4.578
Ch32	Low	2.198	2.997
Ch33	Low	2.689	2.675
Ch35	Low	2.803	2.401
Ch36	Low		
Ch37	Low		
Ch7	Mid	2.394	1.145
Ch10	Mid	2.566	1.686
Ch12	Mid	1.966	2.153
Ch13	Mid	2.435	2.631
Ch14	Mid		
Ch15	Mid	1.627	2.301
Ch16	Mid	4.537	3.613
Ch19	Mid	2.231	2.005
Ch24	Mid	2.741	1.819
Ch25	Mid	3.639	2.391
Ch26	Mid	1.745	2.200
Ch29	Mid	1.678	2.675
Ch34	Mid	3.222	2.596
Ch3	High	3.795	2.465
Ch6	High	2.289	2.788
Ch9	High	3.133	2.601
Ch11	High	0.628	3.084
Ch17	High	2.264	2.070
Ch18	High	2.356	2.329
Ch20	High	2.214	0.961
Ch21	High	2.232	1.116
Ch27	High	1.592	2.632
Ch28	High	2.474	1.476
Ch31	High		

Native speakers' individual means of length of eye looks to competitor and target pictures in verbal agreement trials

Participant NO	Target looks in second	Competitor looks in second
NS1	2.657	1.180
NS2	2.580	1.030
NS3	2.715	1.039
NS4	2.498	1.127
NS5	2.260	1.250
NS6	2.306	1.428
NS7	2.195	1.601
NS8	2.891	1.183
NS9	2.457	1.201
NS10	2.404	1.305

Arab participant's individual means of length of eye looks to competitor and target pictures in verbal agreement trials

Participant NO	Proficiency Level	Target looks in second	Competitor looks in second
A2	Low	2.220	1.254
A5	Low	2.421	2.790
A6	Low	2.949	2.108
A8	Low	2.330	1.267
A15	Low		
A21	Low		
A22	Low	3.728	0.806
A24	Low	2.470	1.799
A27	Low	1.478	1.511
A28	Low	2.419	0.272
A32	Low	3.418	2.909
A1	Mid	2.619	3.059
A4	Mid		
A9	Mid	2.893	0.947
A11	Mid	2.448	1.300
A13	Mid	4.377	2.278
A14	Mid	2.580	2.262
A16	Mid	2.304	1.371
A17	Mid	2.220	2.023
A18	Mid	2.596	1.925
A19	Mid	2.213	1.634
A23	Mid	3.004	1.401
A25	Mid	2.200	2.064
A26	Mid	3.176	1.772
A34	Mid	3.497	3.242
A3	High	2.591	1.069
A7	High	2.905	0.973
A10	High	3.152	1.808
A12	High	2.570	1.243
A20	High	2.397	1.283
A29	High	2.117	1.446
A30	High	2.746	1.177
A31	High	2.531	1.229
A33	High	2.088	1.141

Chinese participant's individual means of length of eye looks to competitor and target pictures in verbal agreement trials

Participant NO	Proficiency Level	Target looks in second	Competitor looks in second
Ch1	Low	2.235	2.087
Ch2	Low	2.085	2.709
Ch4	Low	3.670	3.364
Ch5	Low	2.724	1.805
Ch8	Low	2.472	1.276
Ch22	Low	3.181	0.678
Ch23	Low	2.530	2.562
Ch30	Low	2.279	1.844
Ch32	Low	2.908	2.512
Ch33	Low	3.696	1.449
Ch35	Low	2.446	1.749
Ch36	Low		
Ch37	Low		
Ch7	Mid	1.937	1.679
Ch10	Mid	2.718	1.517
Ch12	Mid	2.599	1.575
Ch13	Mid	1.955	1.800
Ch14	Mid	2.680	2.462
Ch15	Mid	2.193	2.939
Ch16	Mid	3.903	3.115
Ch19	Mid	2.894	0.744
Ch24	Mid	2.814	1.579
Ch25	Mid	2.704	2.244
Ch26	Mid	2.754	1.013
Ch29	Mid	2.607	1.492
Ch34	Mid	2.935	1.469
Ch3	High	2.336	3.397
Ch6	High	2.496	1.823
Ch9	High	2.382	1.630
Ch11	High	2.334	0.783
Ch17	High	2.509	1.579
Ch18	High	2.513	1.585
Ch20	High	2.514	1.950
Ch21	High	2.677	1.219
Ch27	High	2.531	1.158
Ch28	High	2.855	1.867
Ch31	High		

F.6. Results from the Computerised Picture Choice Task: Individual Means of Speed of First Look to Target

Native speakers' individual means of speed of first look to target picture in past tense trials

Participant NO	First Look to target means in millisecond
NS1	645
NS2	888
NS3	284
NS4	569
NS5	942
NS6	1081
NS7	851
NS8	473
NS9	679
NS10	540

Arab participants' individual means of speed of first look to target picture in past tense trials

Participant NO	Proficiency Level	First Look to target means in millisecond
A2	Low	1457
A5	Low	886
A6	Low	931
A8	Low	1817
A15	Low	
A21	Low	
A22	Low	828
A24	Low	1082
A27	Low	1170
A28	Low	2021
A32	Low	1023
A1	Mid	1035
A4	Mid	
A9	Mid	765
A11	Mid	339
A13	Mid	1410
A14	Mid	932
A16	Mid	977
A17	Mid	806
A18	Mid	915
A19	Mid	827
A23	Mid	1008
A25	Mid	346
A26	Mid	1226
A34	Mid	1174
A3	High	540
A7	High	467
A10	High	717
A12	High	636
A20	High	1081
A29	High	798
A30	High	1695
A31	High	479
A33	High	549

Chinese participants' individual means of speed of first look to target picture in past tense trials

Participant NO	Proficiency Level	First Look to target means in millisecond
Ch1	Low	1534
Ch2	Low	1601
Ch4	Low	994
Ch5	Low	1626
Ch8	Low	1319
Ch22	Low	2217
Ch23	Low	1804
Ch30	Low	2215
Ch32	Low	1913
Ch33	Low	1434
Ch35	Low	1540
Ch36	Low	
Ch37	Low	
Ch7	Mid	413
Ch10	Mid	575
Ch12	Mid	1061
Ch13	Mid	837
Ch14	Mid	
Ch15	Mid	857
Ch16	Mid	1810
Ch19	Mid	1651
Ch24	Mid	1620
Ch25	Mid	910
Ch26	Mid	1369
Ch29	Mid	1559
Ch34	Mid	855
Ch3	High	1614
Ch6	High	1290
Ch9	High	1953
Ch11	High	1836
Ch17	High	1767
Ch18	High	990
Ch20	High	494
Ch21	High	.
Ch27	High	1969
Ch28	High	912
Ch31	High	

Native speakers' individual means of speed of first look to target picture in verbal agreement trials

Participant NO	First Look to target means in millisecond
NS1	605
NS2	633
NS3	299
NS4	385
NS5	847
NS6	558
NS7	690
NS8	884
NS9	407
NS10	586

Arab participants' individual means of speed of first look to target picture in verbal agreement trials

Participant NO	Proficiency Level	First Look to target means in millisecond
A2	Low	1183
A5	Low	1103
A6	Low	800
A8	Low	925
A15	Low	
A21	Low	
A22	Low	1270
A24	Low	1592
A27	Low	1240
A28	Low	950
A32	Low	926
A1	Mid	636
A4	Mid	
A9	Mid	683
A11	Mid	942
A13	Mid	1317
A14	Mid	771
A16	Mid	2356
A17	Mid	1262
A18	Mid	692
A19	Mid	635
A23	Mid	729
A25	Mid	851
A26	Mid	888
A34	Mid	379
A3	High	777
A7	High	665
A10	High	690
A12	High	377
A20	High	826
A29	High	879
A30	High	1104
A31	High	642
A33	High	714

Chinese participants' individual means of speed of first look to target picture in verbal agreement trials

Participant NO	Proficiency Level	First Look to target means in millisecond
Ch1	Low	1440
Ch2	Low	1015
Ch4	Low	1007
Ch5	Low	1100
Ch8	Low	1147
Ch22	Low	1100
Ch23	Low	1159
Ch30	Low	1532
Ch32	Low	1323
Ch33	Low	1300
Ch35	Low	1022
Ch36	Low	
Ch37	Low	
Ch7	Mid	537
Ch10	Mid	366
Ch12	Mid	977
Ch13	Mid	1172
Ch14	Mid	1186
Ch15	Mid	906
Ch16	Mid	652
Ch19	Mid	616
Ch24	Mid	578
Ch25	Mid	628
Ch26	Mid	918
Ch29	Mid	1443
Ch34	Mid	1439
Ch3	High	1119
Ch6	High	800
Ch9	High	791
Ch11	High	955
Ch17	High	1448
Ch18	High	1798
Ch20	High	708
Ch21	High	.
Ch27	High	1039
Ch28	High	757
Ch31	High	