MORPHOLOGICAL VARIABILITY IN SECOND LANGUAGE ARABIC

By

Baraa A. Rajab
A Dissertation
Submitted to the
Graduate Faculty
of
George Mason University
in Partial Fulfillment of
The Requirements for the Degree
of
Doctor of Philosophy
Linguistics

Committee:

___________________________________________ Director

___________________________________________

___________________________________________ Department Chairperson

___________________________________________ Program Director

___________________________________________ Dean, College of Humanities
and Social Sciences

Date: ________________________________ Fall Semester 2017
George Mason University
Fairfax, VA
Morphological Variability in Second Language Arabic

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at George Mason University

By

Baraa Rajab
Master of Arts
George Mason University, 2010
Master of Arts
King Abdul Aziz University, 2007

Director: Steven Weinberger, Associate Professor
Department of English, Program in Linguistics

Fall Semester 2017
George Mason University
Fairfax, VA
Dedication

This is dedicated to my parents and siblings.
Acknowledgements

I would like to thank my family, friends, colleagues and professors who helped make this possible. My parents, who have supported me and stood by me. Thank you my sisters for keeping up with “I’m busy” and standing by me every step of the way. Dr. Jenny Culberston, without you I would not be where I am today. Thank you for your guidance, support and invaluable feedback through this process. Dr. Steven Weinberger and Dr. Doug Wulf, your continuous encouragement and inspiration and making my graduate school experience run smoothly. Dr. Greenberg thank you for your support. A special thanks to Dr. Charlie Jones. I wish you were with me here today to share this moment, but you will always be in my heart.
# Table of Contents

List of Tables ...................................................................................................................... vii
List of Figures ......................................................................................................................... ix
List of Abbreviations ................................................................................................................. x
Abstract .................................................................................................................................. xi

## Chapter 1. Introduction

1.1 MSIH and Comprehension ................................................................. 1
  1.1.1 Acquisition of Arabic Morphosyntax ........................................... 3
1.2 Linguistic Background ........................................................................ 4
  1.2.1 Arabic Gender and Number ....................................................... 4
  1.2.2 Adjective-Noun Agreement in Modern Standard Arabic .......... 7
  1.2.3 Direct Objects Clitics in Arabic .................................................. 8
  1.2.4 Direct Object Clitic Agreement in MSA ..................................... 10

## Chapter 2. Comprehension and Production of Gender and Number Agreement in L2 Arabic NPs

2.1 Experiment 1 ....................................................................................... 13
  2.1.1 Participants ............................................................................... 13
  2.1.2 Elicited Production ................................................................. 13
  2.1.3 Comprehension ....................................................................... 16
  2.2 Results ............................................................................................ 18
    2.2.1 Elicited Production Results .................................................... 18
    2.2.2 Comprehension Results ......................................................... 20
  2.3 Discussion ........................................................................................ 22
  2.4 Conclusion ...................................................................................... 24

## Chapter 3. Morphological Variability in L2 Arabic NPs

3.1 Experiment 2 ....................................................................................... 26
  3.1.1 Participants ............................................................................... 26
  3.1.2 Elicited Production ................................................................. 26
  3.1.3 Comprehension ....................................................................... 29
  3.2 Results ............................................................................................ 31
    3.2.1 Production Results ............................................................... 31
    3.2.2 Comprehension Results ......................................................... 33
  3.3 General Discussion ......................................................................... 36
  3.4 Conclusion ...................................................................................... 38

## Chapter 4. Morphological Variability in L2 Arabic Direct Object Clitic Agreement

4.1 DO Clitics in Modern Standard Arabic ......................................................... 40
  4.1.1 Direct Object Clitic Agreement in Modern Standard Arabic .... 40
4.2 Experiment 3 ........................................................................................................... 74
  4.2.1 Participants ........................................................................................................ 74
  4.2.2 Elicited Production ......................................................................................... 74
  4.2.3 Comprehension ............................................................................................... 77
4.3 Results .................................................................................................................... 79
  4.3.1 Production Results ........................................................................................ 80
  4.3.2 Comprehension Results ................................................................................ 93
4.4 General Discussion ............................................................................................... 105
4.5 Conclusion ............................................................................................................. 108
Chapter 5. Conclusion ................................................................................................. 110
  5.1 Summary of Major Findings ............................................................................. 110
  5.2 Questions for Future Research ......................................................................... 112
Appendix A: Demographic Questions ....................................................................... 114
Appendix B: Vocabulary Set 1 and 2 for Production and Comprehension in NPs
  Experiment .............................................................................................................. 115
Appendix C: Vocabulary Set 1 and 2 for Production and Comprehension in DO Clitics
  Experiment .............................................................................................................. 117
References ................................................................................................................ 119
List of Tables

Table 1.1 Schematic picture of adjectival agreement paradigm in Modern Standard Arabic .......................................................... 11
Table 1.2. 3rd Person Direct Object Clitics in Modern Standard Arabic ........................................ 12
Table 2.1. Elicited Production Task Items by Gender and Number in Pilot Study .......... 18
Table 2.2. Tentative Adjectives for Production Task by Gender and Number in Pilot Study .......................................................... 19
Table 2.3. Mean % accuracy in NPs during the production task by proficiency group and gender feature .......................................................... 23
Table 2.4. Mean % accuracy in NPs during the production task, by proficiency group and number feature .......................................................... 24
Table 2.5. Mean % accuracy of number in NPs by native speakers during the production task ........................................................................ 24
Table 2.6. Mean % production of dual target nouns and adjectives with each type of number agreement in the native speaker group ........................................................................ 25
Table 2.7. Mean % adjective number for dual target noun produced, for each group ........................................................................ 25
Table 2.8. Mean % accuracy on the comprehension task, by proficiency group and feature ........................................................................ 26
Table 2.9. Average % production of adjectives with each number agreement type with dual animate nouns by proficiency group ........................................................................ 29
Table 2.10. Average % production of adjective with each number agreement type number with dual inanimate nouns by proficiency group ........................................................................ 30
Table 3.1. Sample Elicited Production Task Items by Gender and Number from Vocabulary Set 1 ........................................................................ 38
Table 3.2. Sample Adjectives by Gender and Number ........................................................................ 38
Table 3.3. Schematic picture of adjectival agreement paradigm in Modern Standard Arabic ........................................................................ 54
Table 3.4. Adjective Gender and Number Response when Target Noun Is Dual Masculine ........................................................................ 54
Table 3.5. Adjective Gender and Number Response when Target Noun Is Dual Feminine ........................................................................ 54
Table 3.6. Number errors for dual nouns by proficiency ........................................................................ 64
Table 3.7. Number errors for plural nouns by proficiency ........................................................................ 65
Table 3.8. Number substitution for non-human plural nouns by proficiency ........................................................................ 65
Table 4.1. Direct Object Clitics in Modern Standard Arabic ........................................................................ 71
Table 4.2 Direct Object Clitics in Modern Standard Arabic ........................................................................ 75
Table 4.3 Schematic picture of DO clitic agreement paradigm in Modern Standard Arabic .............................................................. 83
Table 4.4. Response Clitic Gender and Number for Non-Human Masculine Plural Target Nouns for LP.................................................................................................................. 87
Table 4.5. Response Clitic Gender and Number for Non-Human Masculine Plural Target Nouns for HP .......................................................... 87
Table 4.6. Clitic Number Response when Target Noun Is Dual for LP Group........... 88
Table 4.7. Clitic Number Response when Target Noun Is Dual for HP Group .......... 88
Table 4.8. Average Production of Clitic Number with Dual Human Nouns by Proficiency .................................................................................. 92
Table 4.9. Average Production of Clitic Number with Dual Non-Human Nouns by Proficiency .................................................................................. 92
Table 4.10. Clitic Number Choice when Target Noun Is Non-Human Masc Pl. for LP Group .................................................................................. 99
Table 4.11. Clitic Number Choice when Target Noun Is Non-Human Masc Pl. for HP Group .................................................................................. 100
Table 4.12. Clitic Number Response when Target Noun Is Dual for LP Group........ 101
Table 4.13. Clitic Number Response when Target Noun Is Dual for HP Group ........ 102
Table 4.14. Average Clitic Choice on Comprehension Task with Human Dual Nouns by Proficiency ................................................................. 104
Table 4.15. Average Clitic Choice on Comprehension Task with Non-Human Dual Nouns by Proficiency ................................................................. 104
### List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>Pilot Study Comprehension task: Sample test item</td>
<td>22</td>
</tr>
<tr>
<td>3.1</td>
<td>Sample test item from the production task. Example: (t’abeeban kabeeran) “two old doctors”</td>
<td>40</td>
</tr>
<tr>
<td>3.2</td>
<td>Sample test item from the comprehension task. hadaa mozareš (“This is a farmer”)</td>
<td>44</td>
</tr>
<tr>
<td>3.3</td>
<td>Mean % accuracy of adjective inflection on the production task, by proficiency and gender across proficiency groups</td>
<td>46</td>
</tr>
<tr>
<td>3.4</td>
<td>Mean % accuracy of adjective inflection on the production task, by proficiency and number across proficiency groups</td>
<td>48</td>
</tr>
<tr>
<td>3.5</td>
<td>Mean % correct for singular, dual and plural adjective inflection for LP</td>
<td>50</td>
</tr>
<tr>
<td>3.6</td>
<td>Mean % correct for singular, dual and plural adjective inflection for HP</td>
<td>51</td>
</tr>
<tr>
<td>3.7</td>
<td>Mean % Accuracy during the comprehension task, by proficiency and gender</td>
<td>60</td>
</tr>
<tr>
<td>3.8</td>
<td>Mean % accuracy of noun inflection on the comprehension task by proficiency and number</td>
<td>61</td>
</tr>
<tr>
<td>3.9</td>
<td>Mean % accuracy on noun inflection by number and animacy for LP</td>
<td>62</td>
</tr>
<tr>
<td>3.10</td>
<td>Mean % accuracy on noun inflection by number and animacy for HP</td>
<td>63</td>
</tr>
<tr>
<td>4.1</td>
<td>Sample production test trial. Example: The boy is pulling him (huwa yashabhu)</td>
<td>77</td>
</tr>
<tr>
<td>4.2</td>
<td>Sample test item from the comprehension task: Howa yohebo-hu (“He Loves him”)</td>
<td>79</td>
</tr>
<tr>
<td>4.3</td>
<td>Mean % accuracy of clitic production by proficiency and gender</td>
<td>81</td>
</tr>
<tr>
<td>4.4</td>
<td>Mean % accuracy of clitics on the production task, by proficiency and number</td>
<td>82</td>
</tr>
<tr>
<td>4.5</td>
<td>Mean % correct for singular, dual and plural clitic production for LP</td>
<td>84</td>
</tr>
<tr>
<td>4.6</td>
<td>Mean % correct for singular, dual and plural clitic production for HP</td>
<td>85</td>
</tr>
<tr>
<td>4.7</td>
<td>Mean % accuracy of clitic inflection on the comprehension task by proficiency and gender</td>
<td>94</td>
</tr>
<tr>
<td>4.8</td>
<td>Mean % accuracy of clitic inflection on the comprehension task by proficiency and number</td>
<td>95</td>
</tr>
<tr>
<td>4.9</td>
<td>Mean % correct clitic choice for human vs. non-human target nouns for LP</td>
<td>96</td>
</tr>
<tr>
<td>4.10</td>
<td>Mean % correct clitic choice for human vs. non-human target nouns for HP</td>
<td>98</td>
</tr>
</tbody>
</table>
**List of Abbreviations**

1st ........................................................................................................... 1st Person  
2nd ........................................................................................................... 2nd Person  
3rd ........................................................................................................... 3rd Person  
ADJ ........................................................................................................... Adjective  
DO ........................................................................................................... Direct Object  
f(em) ....................................................................................................... fem  
FFFH ....................................................................................................... Failed Functional Feature Hypothesis  
IL ........................................................................................................... Interlanguage  
L1 ........................................................................................................... First Language  
L2 ........................................................................................................... Second Language  
L2er ....................................................................................................... Second Language Learner  
m(asc) ..................................................................................................... masculine  
MSA ....................................................................................................... Modern Standard Arabic  
MSIH ..................................................................................................... Missing Surface Inflection Hypothesis  
MUSH ..................................................................................................... Morphological Underspecification Hypothesis  
N ........................................................................................................... Noun  
O ........................................................................................................... Object  
pl .............................................................................................................. plural  
S .............................................................................................................. Subject  
sing ......................................................................................................... singular  
SLA ....................................................................................................... Second Language Acquisition  
V ............................................................................................................. Verb
Abstract

MORPHOLOGICAL VARIABILITY IN SECOND LANGUAGE ARABIC

Baraa Rajab, Ph.D.

George Mason University, 2017

Dissertation Director: Dr. Steven Weinberger

A prominent theory of variability in L2 learners’ use of inflectional morphology is the Missing Surface Inflection Hypothesis (MSIH). The MSIH argues that morphological variability results from ‘performance limitations’ - particularly when a learner is under communication pressure (Prevost&White, 2000). Such pressure is expected to be highest during production, and lower during receptive tasks, and thus many studies of L2 variability have focused exclusively on production. Recent work has shown that variable comprehension of agreement also occurs in second language learners (L2ers) (McCarthy, 2008). This suggests that receptive tasks can help adjudicate among theories of morphological variability and into the mechanisms responsible. McCarthy (2007) also suggests that L2ers resort to the underspecified form when making production or comprehension errors; Morphological Underspecification Hypothesis (MUSH). Nevertheless, studies comparing production and comprehension are few and are restricted mainly to the Romance languages. This dissertation investigates morphological
variability in gender and number agreement in English-speaking L2 learners’ production and comprehension of Arabic NP agreement and DO clitic agreement. The focus on Arabic is motivated by two important gaps: first, Arabic L2 acquisition is relatively understudied; second Arabic has full agreement in gender and number, with a three-way number distinction (singular/dual/plural) in NPs and DO clitics. The acquisition of dual by L2ers not encoding this distinction is not well understood. Here I use L1 speakers of English; a language that does not encode number or gender agreement among nominal constituents. Data from 3 cross-sectional experiments at 2 proficiency levels were collected to test the predictions of the MSIH (Prevost and White, 2000) and MUSH (McCarthy, 2007). Results of the experiments suggest (1) Morphological variability is a persistent problem for L2ers even at advanced proficiency levels; (2) Morphological Variability extends to comprehension, however it decreases in comprehension; (3) Animacy plays a role in the acquisition of agreement; where human targets were numerically higher in agreement accuracy than non-human targets; (4) the use of feminine adjectives in masculine contexts suggesting that feminine is the default for errors in Arabic. On the other hand, since variability clearly decreased in comprehension compared to production these results confirm predictions of the MSIH—decreasing communication pressure decreases variability.
Chapter 1. Introduction

Morphological variability is a well-attested finding in second language learners (L2ers): overt morphemes are at some stage inconsistently produced. Research has shown that variability extends to different proficiency levels, different languages being acquired and across different domains (Alhawary, 2005; Franceschina, 2001; Lardiere, 1998 and White, 2003). A number of previous studies in second language acquisition have investigated the extent to which morphological variability in the use of inflection reflects underlying grammatical knowledge. Some researchers propose that the absence or variable use of inflection in second language production is due to syntactic or representational impairment in the interlanguage (IL) grammar, whether permanent (Meisel, 1991; Hawkins and Chan, 1997) or temporary (Vainikka and Young-Scholten, 1996). However, an alternative view—the Missing Surface Inflection Hypothesis (MSIH)—argues that variability of inflectional morphology results instead from the learner’s inability to access the fully specified form due to ‘blocking’, ‘communication pressure’ (Prevost & White, 2000), or ‘mapping problems’ (Lardiere, 1998). Such issues are typically thought to arise most prominently during production, and thus many studies of variability have focused on the productions of L2 learners. Importantly, recent work has shown that variable comprehension of agreement can also occur in second language learners, suggesting that tasks probing receptive grammar can provide additional insight
into the mechanism responsible for agreement errors. McCarthy (2008), for example, showed that patterns of errors were qualitatively similar across comprehension and production tasks assessing Spanish L2 learners’ knowledge of agreement. She argues that this continuity across tasks suggests that morphological variability is due to underlying grammatical knowledge and not performance issues alone. Despite the importance of this line of research for evaluating theories of L2 morphological variability, studies explicitly manipulating communication pressure or other task effects remain relatively rare.

This dissertation builds on the work of McCarthy (2007; 2008) and others by examining English-speaking L2 learners’ production and comprehension of Arabic morphological inflectional agreement in the nominal domain. I focus on Arabic for two main reasons: first it is a relatively understudied language in terms of L2 acquisition, and second it has full agreement of elements within the noun phrases, not only in gender but number, where it maintains a three-way distinction—singular/dual/plural. To date, the acquisition of dual by speakers of a language which does not encode this distinction has not been systematically investigated. Section 1.2 of this chapter provides some additional background on studies of morphological variability in production and comprehension relevant to this dissertation, and L2 acquisition of Arabic. It also discusses the major research questions of this dissertation. I then provide a linguistic background on Modern Standard Arabic Noun Phrases and direct object clitic agreement in Section 1.3.
Chapter Two reports a pilot study addressing the acquisition of features of gender and number in the nominal domain specifically in noun phrases that consist of a head noun and an attributive adjective in second language Arabic.

Chapter Three and Four consist of experimental tests in production and comprehension across domains. Chapter Three looks at morphological variability in the nominal domain specifically N + ADJ combinations with a focus on animacy features. Chapter Four considers variability in the verbal domain specifically gender and number agreement in direct object (DO) clitics.

Chapter Five summarizes the major findings of this dissertation, and suggests new directions for future research.

1.1 MSIH and Comprehension

In the domain of noun phrase agreement, several recent studies have focused both on distinctions between production and receptive tasks and the particular types of errors that L2 learners tend to make. These studies have generally reported systematicity in errors of gender and number assignment; typically, masculine gender and singular number are overused across different tasks in a number of languages (Ellis et al., 2012; Hopp, 2013; McCarthy, 2008; Montrul et al., 2008; Santoro, 2012 and White et al., 2004). For example, Santoro (2012) examined acquisition of morphological features of gender and number agreement in L2 Italian determiners, adjectives and pronouns. Adult native speakers of English, divided into two proficiency groups (high beginners and high intermediate), were tested in oral and written production tasks. Overall, results showed that accuracy levels were higher for masculine and singular targets than for feminine or
plural ones. This recurring finding has been argued to reflect the assignment of default features (Alacorn, 2010; Martinez-Gibson, 2011 and Renaud 2010; 2011). More importantly, Santoro (2012) compared results from oral production and written production, and found higher accuracy rates in the written task (in both proficiency groups) compared to the oral task. Santoro (2012) interprets these findings as support for the MSIH. In other words, morphological variability occurs predominantly in production; problems in the mapping between morphosyntax and surface phonological representation should occur more often when the pressure to communicate is most strong.

Although establishing a connection between performance limitations (or communication pressure) and variability in the realization of morphology is crucial for evaluating the MSIH. The vast majority of studies on morphological variability have looked only at production (written and/or oral). If, as the MSIH argues, production tasks mask learners’ underlying competence, then comprehension should be relatively spared. To date, a few studies have investigated this directly. Both White et al. (2004) and McCarthy (2008) tested the predictions of the MSIH on data from acquisition of gender and number agreement in L2 Spanish. The results of these two studies are consistent with one another–finding clear parallels between error patterns in production and comprehension–so here I discuss only McCarthy (2008) in detail. In this study, native English L2 speakers of Spanish, divided into 2 proficiency levels (intermediate and advanced), were tested on direct object (DO) clitics and adjectival agreement. In production, variability in gender and number agreement with adjectives was found even in advanced L2 learners. Participants tended to use singular agreement in plural contexts,
and masculine gender in feminine contexts. Importantly, in comprehension, results were very similar; participants used masculine gender in feminine contexts, and singular number in plural contexts even in advanced L2 grammars. McCarthy (2008) argued that since similar variability in L2ers ability to use agreement was found in both production and comprehension, the explanation cannot be purely attributable to performance. Rather, she suggests that eliminating communication pressure does not eliminate variability. Moreover, since learners’ errors were similar in both tasks (overuse of masculine gender and singular number), learners may actually be failing to fully represent the relevant features, especially in intermediate proficiency groups. This missing representational knowledge would then lead to errors in both task types, contrary to the predictions of the MSIH. McCarthy (2008) suggests that these results are evidence for a morphological deficit as observed from the “emergence of underspecified default morphology” (p.480) or what she calls the MUSH – Morphological Underspecification Hypothesis (For a detailed discussion, see McCarthy, 2007). In addition, recent studies on Morphological variability in second language acquisition have shown that animacy and semantic gender play a role in the acquisition of agreement. Alarcon (2010) found that L2 learners of Spanish are sensitive to linguistic cues such as semantic gender and animacy to establish correct agreement features.

The studies mentioned above clearly help to shed some light into the acquisition process of morphological agreement in L2 grammars. However, the question remains whether the source of variability lies in the underlying representations or in limitations of the production system. In particular, while morphological variability appears to be
attested in comprehension, and to be qualitatively similar to that found in production, the evidence remains limited. Additional studies are needed to replicate these findings in a wider variety of languages and features. Further it is not clear if the level of variability differs in interesting ways; if error rates are lower in interpreting agreement features during comprehension than in producing correct agreement, this could be consistent with the MSIH.

1.1.1 Acquisition of Arabic Morphosyntax

Most research on L2 acquisition in general, and morphological variability in particular, are focused on relatively few languages (e.g., Spanish, French, German, and, Italian). Here I aim to extend this line of research by looking at a different language, with a different set of agreement features, Arabic. The acquisition of functional features and categories in Arabic as a second language has not yet been studied extensively. There seems to be a general consensus that word order is unproblematic for Arabic L2 learners (Albrini et al. 2013; Alhawary 2005; 2009, Alduais 2012), similar to what has been observed in other L2s (White et al., 2004). In addition, masculine gender has generally been found to be mastered earlier than feminine (Alhawary 2005; 2009, Alduais 2012). For example, one of the first studies was conducted by Alhawary (2005) who investigates agreement in L2 Arabic NPs. Native English and French speakers, divided into 3 proficiency levels, were tested using a semi-elicited spontaneous production task. All of the participants showed 100% correct Noun-Adjective word order. In gender agreement, however, variability was clearly attested: while all learners showed high accuracy rates
on masculine singular agreement, lower proficiency learners showed significantly more errors in feminine agreement (in line with many others studies).

In a later study, Alhawary (2009) collected longitudinal and cross-sectional data on nominal agreement from native English, French and Japanese speakers learning Arabic. He again found that learners consistently produced correct word order (noun-adjective), and were near ceiling at masculine gender agreement. However, when producing noun-adjective combinations, he found that the participants often supplied the correct feminine morpheme on the noun but not the adjective. In demonstrative-noun combinations, participants would overuse the masculine demonstrative in feminine contexts. Alhawary (2009) concludes that L2 learners of Arabic use the masculine as the default form.

Delayed mastery of feminine relative to masculine morphology is not just observed in L2 acquisition, but in L1 acquisition as well. Omar (1973; 2007) reported the emergence of masculine before feminine forms in Arabic-learning children. Masculine singular has been widely argued to be the default category in Arabic nouns since it is the bare form, with no overt morphology (e.g., Albrini et al., 2013).

Notably, work on acquisition of number in L2 speakers of Arabic is currently missing. Studies on L2 acquisition of number agreement in a number of other languages have found very few errors in singular morphology and over-extension of singular morphology to plurals (Ellis et al., 2012; Hopp, 2013; McCarthy, 2008; Montrul et al., 2008; Santoro, 2012 and White et al., 2004). It remains to be seen whether that pattern extends to Arabic.
This dissertation follows McCarthy (2008), investigating patterns of morphological variability across production and comprehension. I aim to address the following questions in a series of behavioral experiments measuring L2 Arabic speakers’ knowledge of number and gender morphology:

1. What are the general features of L2 speakers’ knowledge of gender and number agreement in Arabic NPs and DO clitics?
2. What are the error patterns of agreement morphology in terms of gender (masculine, feminine) and number (singular, dual, plural) within NPs and DO clitics?
3. Is morphological variability in the production task similar to morphological variability in the comprehension task?
4. What is the role of animacy (human vs. non-human) in the acquisition of Arabic L2 gender and number features?

By addressing these questions, this dissertation will test the predictions of the two major theories of morphological variability: the Missing Surface Inflection Hypothesis MSIH (Prevost & White, 2000) and the Morphological Underspecification Hypothesis MUSH (McCarthy, 2007). To reiterate: the MSIH, errors arise due to “communication pressure”. Therefore, the MSIH predicts that inflection may be dropped when such pressures are high, e.g., during production, but not otherwise, e.g., during comprehension. When inflection is supplied it should be accurate. By contrast, the MUSH predicts underspecified representation of features in learners’ IL grammar. Therefore variability is predicted to surface in both production and comprehension. Since the same
underspecified representations are available across comprehension and production, we should expect that variability is qualitatively similar – underspecified morphemes act as defaults.

1.2 Linguistic Background

The English and Arabic nominal systems differ both in terms of the features expressed and in terms of agreement targets. Unlike English, Arabic gives morphological expression to number and gender features on nouns and direct object clitics, and agreement between nouns and nominal modifiers is much richer in Arabic compared to English. Below, the relevant differences between these systems are briefly highlighted.

1.2.1 Arabic Gender and Number

Arabic has two genders—masculine and feminine—morphologically encoded in nouns, and triggering pronominal, verbal and adjectival agreement. Although there are a number of irregular declension, in the most frequent case masculine is realized with a zero morpheme (kitab-Ø – ‘book’), while feminine is realized with the addition of a suffix –a¹ (sayar-a – ‘car’) (Brustad, Al-Batal, & Tūnisī, 2004).

Arabic has three number distinctions: singular, dual and plural (Ryding 2005). Because dual is reserved for numerosity two, the plural refers to groups of three or more. Singular number is marked by a zero morpheme. The plural is generally marked either by the suffix -uun/-iin for masculine, -aat for feminine or by a change to the noun stem. The dual suffix is –an (kitab-an ‘two books’) for masculine and –tan (sayyara-tan ‘2 cars’) for feminine.

¹ Feminine nouns like ‘shams’ sun and ‘naar’ fire that are not overtly marked were excluded
1.2.2 Adjective-Noun Agreement in Modern Standard Arabic

NPs in Modern Standard Arabic Arabic consisting of a head noun and an attributive adjective involve agreement in gender (masculine or feminine), number (singular, dual or plural), and case\(^2\) (nominative, accusative, or genitive). Plural agreement has an animacy-based distinction between human and non-human entities. Human plurals trigger matching agreement on nominal modifiers and the verb. However, non-human plurals invariably take singular feminine adjectives regardless of the object’s gender (for more on plurals, see Alhawary, 2011). Examples 1–3 below illustrate agreement between nouns and adjectives, and the distinction between human and non-human plurals.

<table>
<thead>
<tr>
<th></th>
<th>Masculine Singular and Dual Nominals</th>
<th></th>
<th>Feminine Singular and Dual Nominals (human/non-human distinction)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(a) maktab–Ø</td>
<td>qadeem–Ø</td>
<td>(c) wazeer–Ø</td>
</tr>
<tr>
<td></td>
<td>desk.Sg.M</td>
<td>old.Sg.M</td>
<td>minister.Sg.M</td>
</tr>
<tr>
<td></td>
<td>‘an old desk’</td>
<td>‘an old minister’</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(b) maktab–an</td>
<td>qadeem–an</td>
<td>(d) wazeer–an</td>
</tr>
<tr>
<td></td>
<td>‘two old desks’</td>
<td>‘two old ministers’</td>
<td></td>
</tr>
</tbody>
</table>

|   | (a) tˤawil–a | qadeem–a | (c) wazeer–a | qadeem–a |
|   | table.Sg.F | old.Sg.F | minister.Sg.F | old.Sg.F |
|   | ‘an old table’ | ‘an old minister (f)’ |   | |
|   | (b) tˤawila–tan | qadeema–tan | (d) wazeer–tan | qadeem–tan |
|   | ‘two old tables’ | ‘two old ministers (f)’ |   | |

---

\(^2\) In the experiment described here, all forms are in nominative case
In summary, the adjective agreement paradigm in Modern Standard Arabic can be visualized schematically as in Table 1.1 below. Each cell indicates the required adjective form given a noun with the features indicated by column and row headers. For example, a masculine, human, plural noun triggers masculine singular adjectival agreement. On the other hand, a masculine, non-human, plural noun, triggers feminine singular adjectival agreement. Grayed cells indicate forms were gender/number distinctions are neutralized.

**Table 1.1 Schematic picture of adjectival agreement paradigm in Modern Standard Arabic**

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
<td>Non-human</td>
</tr>
<tr>
<td>Singular</td>
<td>M.Sg.</td>
<td>M.Sg.</td>
</tr>
<tr>
<td>Dual</td>
<td>M.Dual</td>
<td>M.Dual</td>
</tr>
<tr>
<td>Plural</td>
<td>M.Pl.</td>
<td>[F.Sg]</td>
</tr>
</tbody>
</table>

### 1.2.3 Direct Objects Clitics in Arabic

Direct Object (DO) pronouns in Arabic are realized as clitics, written as suffixes attached to the end of verbs. Like all pronouns in Arabic, DO clitics are marked for person, gender and number. For purposes of this dissertation, only 3rd person is used. Table 1.2 below shows the DO clitics available in Modern Standard Arabic.
Table 1.2. 3rd Person Direct Object Clitics in Modern Standard Arabic

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>Human</td>
<td>-hu</td>
</tr>
<tr>
<td></td>
<td>Non Human</td>
<td>-ha</td>
</tr>
<tr>
<td>Dual</td>
<td>Human</td>
<td>-huma</td>
</tr>
<tr>
<td></td>
<td>Non Human</td>
<td>-huma</td>
</tr>
<tr>
<td>Plural</td>
<td>Human</td>
<td>-hum</td>
</tr>
<tr>
<td></td>
<td>Non Human</td>
<td>-ha</td>
</tr>
</tbody>
</table>

1.2.4 Direct Object Clitic Agreement in MSA

DOs in Arabic involve agreement in gender (masculine or feminine), number (singular, dual or plural), and person\(^3\) (nominative, accusative, or genitive) with the noun it is referring to. As with noun+adjective constructions, plural agreement in DO clitics exhibits an animacy-based distinction between human and non-human entities. Human plurals trigger matching agreement. However, non-human plurals invariably take singular feminine clitics regardless of the object’s gender. Examples 4–6 below illustrate agreement in DO clitics, and the distinction between human and non-humans.

4. Masculine Singular and Dual DO Clitics (Human/Non-Human)
   (a) al-walad \(\text{d}^\text{a}\text{rab}\) at\(^\text{a}\text{alib}\)
   the-boy hit the student.Sg.M
      al-walad \(\text{d}^\text{a}\text{rabu-hu}\)
      the-boy hit – CL.Sg.M
   (b) al-walad \(\text{d}^\text{a}\text{rab}\) al-kalb
      the-boy hit the dog.Sg.M

\(^3\) In the experiments, all forms are in 3\(^{\text{rd}}\) person
al-walad   d'arabu-hu
the-boy    hit – CL.Sg.M

(c) al-walad   d'arab   at'aalibayn
the-boy    hit          the student.Dual.M

al-walad    d'arabu-huma
the-boy    hit – CL.Dual.M

(d) al-walad   d'arab   al-kalbayn
the-boy    hit          the dog.Dual.M

al-walad    d'arabu-huma
the-boy    hit – CL.Dual.M

5. Feminine Singular and Dual DO Clitics (Human/Non-Human)
(a) al-walad    d'arab    at'aaliba
the-boy    hit          the student.Sg.F

al-walad    d'araba-ha
the-boy    hit – CL.Sg.F

(b) al-walad    d'arab    al-qitt'a
the-boy    hit          the cat.Sg.F

al-walad    d'araba-ha
the-boy    hit – CL.Sg.F

(c) al-walad    d'arab    at'aalibatayn
the-boy    hit          the student.Dual.F

al-walad    d'arabu-huma
the-boy    hit – CL.Dual.F

(d) al-walad    d'arab    al-qitt'a
the-boy    hit          the cat.Sg.F

al-walad    d'araba-huma
the-boy    hit – CL.Dual.F
This chapter has discussed previous work on morphological variability in second language acquisition in various L2 languages. It also explored relevant studies in morphological variability in L2 Arabic. In addition, this chapter discussed the two prominent theories in L2 research that will be tested in this dissertation; MSIH Prevost and White, 2000) and MUSH (McCarthy, 2007). The following chapter explores a pilot study i.e. experiment 1, that tests gender and number agreement in L2 Arabic NPs.
Chapter 2. Comprehension and Production of Gender and Number Agreement in L2 Arabic NPs

This study addresses the acquisition of features of gender and number in the nominal domain specifically in noun phrases that consist of a head noun and an attributive adjective. The focus will be on the acquisition of agreement morphology on the noun and adjective in terms of gender (masculine/feminine) and number (singular/dual) by native English speakers. The choice of English as the L1 is motivated by the extent of differences between the English and Arabic systems; English has no grammatical gender agreement, in fact no noun-adjective agreement at all, and no dual number. Although previous studies of gender agreement has mainly focused on the issue if the gender feature is acquirable in the L2 or not especially if the L1 does not have the feature gender (Hawkins, 2001; Bruhn de Garavito and White, 2002; White et al., 2004; Franceschina, 2001, Sabourin, 2003). Findings in general support the claim if the L1 has the feature gender it will facilitate the acquisition of gender in the L2. However, it is not a necessary nor a sufficient factor in L2 success, as L2 gender has reached native-like performance when the feature is not in the L1 (White et al, 2004) and on the other hand is problematic when the L1 has the feature gender (Bruhn de Garavito & White, 2002). Based on previous findings, this system should present a challenge to English learners, and this pairing of L1 and L2 features has not yet been investigated systematically. The
first task addresses learners’ production of gender and number agreement on attributive adjectives. The second task investigates learners’ ability to use these same morphological cues in comprehension.

This study uses a set of behavioral experiments to measure L2 Arabic speakers’ knowledge of number and gender morphology in the nominal domain. The study seeks to answer the following questions:

1. What are the general features of L2 speakers’ knowledge of gender and number agreement in Arabic NPs?
2. What are the error patterns of agreement morphology in terms of gender (masculine, feminine) and number (singular, dual) within NPs?
3. Is morphological variability in the production task similar to morphological variability in the comprehension task?

Previous research on L2 morphological variability points to a number of expected patterns: (1) development of gender-number agreement in L2 Arabic NPs will progress gradually with increased exposure (2) L2ers will have less trouble with zero-marked masculine-singular agreement (the default; Albirini et al. 2013; Omar, 2007), (3) L2ers will over-extend masculine and singular to feminine and dual contexts. In addition, following McCarthy (2008) and White et al. (2004) we expect to see variability in comprehension as well as production. If the level of variability as well as error rates were similar across tasks, this would support McCarthy’s (2008) claim that variability is not just a performance issue, but likely extends to learners’ underlying morphological
competence. On the other hand, if error rates decrease in the comprehension task, or error patterns are distinct, then this would arguably be consistent with the MSIH.

2.1 Experiment 1

The aim of the study was to provide a quantitative and qualitative comparison of errors in comprehension and production tasks, and to investigate the development of knowledge of the dual across proficiency levels.

2.1.1 Participants

20 speakers of L2 Arabic were recruited from Arabic language courses at a major university in Northern Virginia. All were native speakers of English with no exposure to Arabic prior to enrolling in classes at the university. Following Alhawary (2005, 2009), participants were grouped according to their placement in the Arabic program at this institution, and their length of exposure to Arabic. Group 1 (low proficiency) were all in their 2nd year of the Arabic program, and had received between 135-180 hours of formal instruction in Arabic. Group 2 (high proficiency) were in their 3rd or 4th year of the program and had received more than 225 hours of instruction. A control group of 10 native speakers of Arabic (excluding heritage speakers) was also tested. All members of the control group spoke the same variety of Arabic (Najdi), spoken in the mid-region of Saudi Arabia.

2.1.2 Elicited Production

In the first task, L2 Arabic speakers were shown pictures and were asked to produce noun-adjective combinations. The goal of the task was to examine gender and number agreement in production.
Materials

Materials for the elicited production task include 12 colored pictures each intended to elicit the production of a singular, dual or plural noun plus an adjective. Of crucial interest here were the singular and dual trials (as previous work has focused on plural). Therefore critical test items were 4 singular (2 masculine, 2 feminine) and 4 dual (2 masculine, 2 feminine) pictures. The remaining 4 items were plural and involved a different set of objects.

below provides the complete list of test items. Table 2.2 provides a list of adjectives which the pictures were meant to elicit. These adjectives agree with the head noun in gender and number as shown.

Table 2.1. Elicited Production Task Items by Gender and Number in Pilot Study

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>kitab ‘book’</td>
<td>ṣāḥiba ‘family’</td>
</tr>
<tr>
<td></td>
<td>bayt ‘house’</td>
<td>qit’a ‘cat’</td>
</tr>
<tr>
<td>Dual</td>
<td>qalaman ‘2 pens’</td>
<td>sayaratun ‘2 cars’</td>
</tr>
<tr>
<td></td>
<td>kalban ‘2 dogs’</td>
<td>tawilatan ‘2 tables’</td>
</tr>
<tr>
<td>Plural</td>
<td>karasi ‘chairs’</td>
<td>majallat ‘magazines’</td>
</tr>
<tr>
<td></td>
<td>dafater ‘notebooks’</td>
<td>ashjar ‘trees’</td>
</tr>
</tbody>
</table>
Table 2.2. Tentative Adjectives for Production Task by Gender and Number in Pilot Study

<table>
<thead>
<tr>
<th></th>
<th><strong>Masculine</strong></th>
<th><strong>Feminine</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td>sayeer ‘small’</td>
<td>sayeera ‘small’</td>
</tr>
<tr>
<td></td>
<td>kabeer ‘big’</td>
<td>kabeera ‘big’</td>
</tr>
<tr>
<td></td>
<td>jameel ‘nice/handsome’</td>
<td>jameela ‘nice/beautiful’</td>
</tr>
<tr>
<td></td>
<td>yali ‘expensive’</td>
<td>yalia ‘expensive’</td>
</tr>
<tr>
<td></td>
<td>raxees ‘cheap’</td>
<td>raxeesa ‘cheap’</td>
</tr>
<tr>
<td></td>
<td>qadeem ‘old’</td>
<td>qadeema ‘old’</td>
</tr>
<tr>
<td><strong>Dual</strong></td>
<td>sayeeran ‘small’</td>
<td>sayeeraatan ‘small’</td>
</tr>
<tr>
<td></td>
<td>kabeeran ‘big’</td>
<td>kabeeratan ‘big’</td>
</tr>
<tr>
<td></td>
<td>jameelan ‘nice/handsome’</td>
<td>jameelatan ‘nice/beautiful’</td>
</tr>
<tr>
<td></td>
<td>yaliayan ‘expensive’</td>
<td>yaliatan ‘expensive’</td>
</tr>
<tr>
<td></td>
<td>raxeesan ‘cheap’</td>
<td>raxeesanatan ‘cheap’</td>
</tr>
<tr>
<td></td>
<td>qadeeman ‘old’</td>
<td>qadeematan ‘old’</td>
</tr>
</tbody>
</table>

For the plural items, they are all non-human plurals. As mentioned previously in section 1.2.2, non-human plurals take singular feminine adjectives as in sentence 3 repeated below in sentence 7.

7. Plural Nominals (illustrating human/non-human distinction)
   (a) makatib desk.Pl.M qadeem–a old.Sg.F
   ‘old desks’
   (c) wuzaraa? minister.Pl.M old.Pl.M
   qudama?
   ‘old ministers’
   (b) tˤawila–at table.Pl.F qadeem–a old.Sg.F
   ‘old tables’
   (d) wazeera–at minister.Pl.F old.Pl.F
   wazaraa? ‘old ministers (F)’

Procedure

Participants were first asked demographic questions prior to administering the task. A sample of the questions are below (for a full list, refer to Appendix A).

1. What is your native language?

2. Do you speak any other languages natively?

3. As an adult, have you learned any other languages?
4. When did you start learning Arabic?

5. How many Arabic courses have you studied so far?

The task was presented using PsychoPy software (Peirce, 2009). Pictures were presented in randomized order for each participant. The participants were asked to name each object and describe it and were given a practice trial to ensure they understood the task. Each picture was displayed for 60 seconds unless the participant clicked on the ‘space’ bar to move onto the next trial. In their descriptions of the objects, participants used NPs containing adjectives (describing size, shape, color, depth and length) and the head noun. Each session with each participant was audio recorded using Audacity Version 2.0 (2013), transcribed and coded by the experimenter.

2.1.3 Comprehension

The second task tested L2 Arabic speakers’ comprehension of gender and number of agreement in a picture identification task. This task tests the learners’ ability to identify the correct picture from the number and gender features in a noun phrase.

Materials

Test materials for the comprehension task consisted of 12 sentences each with 3 picture choices—a matching (correct) picture, a picture with the incorrect gender, and a picture with the incorrect number. In order to manipulate gender, all the pictures in the comprehension task were human. As in the production task, 4 trials featured a masculine target (2 singular and 2 dual) and 4 a feminine target (2 singular and 2 dual). The remaining 4 targets were plural (2 masculine and 2 feminine). The noun phrases were embedded in Equational (i.e. verbless) sentences consisting of a determiner and a noun,
and show gender (masculine/feminine) and number (singular, dual and plural) agreement. It is important to note that the plural determiner ‘haʔulaa?’ (‘these’) does not have gender distinctions. Examples 8, 9 and 10 show the agreement paradigm in Equational sentences in singular, dual and plural respectively. These sentences were produced by native Arabic speaker (the experimenter) and recorded for playback during the experiment.

8. (a) haðaa wazeer
   this.Sg.M minister.Sg.M
   “This is a minister”
  (b) haðihee wazeer – a
   this.Sg.F minister.Sg.F
   “This is a minister”

9. (a) haðaan wazeer – an
   this.Dual.M minister. Dual.M
   “These are two ministers”
  (b) hataan wazeera – tan
   this. Dual.F minister.Dual.F
   “These are two ministers”

10. (a) haʔulaa? wuzaraa?
     these.Pl minister.Pl.M
     “These are ministers”
 (b) haʔulaa? wazeera – at
     these.Pl minister.Pl.F
     “These are ministers”

Procedure

Participants proceeding to the comprehension task directly after the production task. On each trial 3 pictures were displayed and the stimulus audio was played. Participants were asked to choose which of three pictures corresponded to the sentence heard by clicking on the appropriate choice. As described above, in each trial, one picture was the target, one had a gender mismatch, and one had a number mismatch. The first picture on the right in each trial is always singular feminine. A sample test item is
demonstrated in Figure 2.1 below. In Figure 2.1, the stimulus is ‘haða a‘beebeb’ (This is a
doctor) referring to masculine singular, therefore the correct choice would be picture 3.

![Figure 2.1. Pilot Study Comprehension task: Sample test item](image)

2.2 Results

Overall accuracy rates within NPs were analyzed by using a series of two-way
ANOVA as in the R Language and Environment for Statistical Computing (R Core
Development Team 2010). Proficiency is a between subjects factor (low, high and native)
and feature (masculine vs. feminine, singular vs. dual) are within subjects factors.

2.2.1 Elicited Production Results

The production task elicited 240 NPs (containing a head noun and attributive
adjective).\(^4\) The data was coded for gender and number agreement, each analyzed
separately for each group. I first discuss participants’ gender agreement production
accuracy rates.

\(^4\) The overall accuracy for one participant from the LP group was more than 2 standard deviations above
the mean of that group. Therefore, the participant was added to the high proficiency group.
Gender agreement within NPs. Table 2.3 below shows the mean accuracy rates for gender agreement across all proficiency levels. As expected, accuracy rates were high across the board for native speakers (NS). That said, some variability in production of target morphology occurs for all proficiency levels and for both masculine and feminine targets. Variability was highest in the LP group. The analysis of gender agreement revealed a significant effect of proficiency (F=5.116; p<.01). Post hoc analysis showed this effect to be driven by the difference between native and LP groups (p<.001). The main effect of gender was not statistically significant (F=2.19; p<.2), however, the interaction between proficiency and gender was significant (F=8.75; p<.0005). A paired-samples t-test shows the feature gender to be statistically significant for the LP group (t=5.14, p<.001) with higher accuracy rates on masculine than feminine; accuracy in the HP group did not differ across masculine and feminine targets (t=0.79; p<.5).

Table 2.3. Mean % accuracy in NPs during the production task by proficiency group and gender feature

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (all)</td>
<td>70.83</td>
<td>78.41</td>
<td>93.75</td>
</tr>
<tr>
<td>Masculine</td>
<td>91.66</td>
<td>77.27</td>
<td>90</td>
</tr>
<tr>
<td>Feminine</td>
<td>50</td>
<td>79.54</td>
<td>97.5</td>
</tr>
</tbody>
</table>

Number Agreement within NPs. The same 240 NPs from all 3 groups were used to analyze number agreement within NPs. Table 2.4 below summarizes the mean percent accuracy for number agreement within NPs by group. As the results show, dramatic variability again exists in all groups. Singular NPs were relatively unproblematic for both LP and HP proficiency levels as well as for native speakers. Dual NPs, on the other hand,
were apparently difficult for both levels of proficiency; LP learners do not supply any dual NPs (0%) and the HP only supply 12.12% correct dual NPs. In addition, the native group shows surprisingly similar results, producing only 8.3% correct dual NPs.

**Table 2.4. Mean % accuracy in NPs during the production task, by proficiency group and number feature**

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number(All)</td>
<td>50</td>
<td>54.55</td>
<td>52.5</td>
</tr>
<tr>
<td>Singular</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Dual</td>
<td>0</td>
<td>12.12</td>
<td>8.3</td>
</tr>
</tbody>
</table>

The analysis of number revealed no significant main effect of group (F=.041; p<.1). Since the native dual NP production shows low accuracy compared to textbook Arabic targets, it is essential to analyze the natives production with respect to dual NPs separately. Before doing so, it is important to show that native speakers do not show any variability in singular and plural NP agreement with respect to number. Table 2.5 below summarizes the results of the natives’ production with respect to number.

**Table 2.5. Mean % accuracy of number in NPs by native speakers during the production task**

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Native</td>
<td>100</td>
<td>8.3</td>
<td>100</td>
</tr>
</tbody>
</table>

Table 2.6 below summarizes native speakers’ use of singular, dual, and plural inflection on nouns and adjectives *when producing dual targets*. What this shows is that
native speakers are split on whether to produce singular or dual inflection on dual target nouns (45% vs. 42.5% of the time). On the other hand, it is clear that the dominant choice for adjective number is singular. This is true regardless of the number marking used on the noun.

Table 2.6. Mean % production of dual target nouns and adjectives with each type of number agreement in the native speaker group

<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>42.5</td>
<td>45</td>
<td>12.5</td>
</tr>
<tr>
<td>Adjectives</td>
<td>82.5</td>
<td>5</td>
<td>12.5</td>
</tr>
</tbody>
</table>

Comparing production of dual marking on the noun reveals a significant effect of group (F=4.23; p<.03). Post hoc analysis showed this effect to be driven by the difference between native and LP groups (p<.01). The LP group produces 16.66% dual nouns in dual contexts, HP 34.09% and natives 45%. No other group differences were significant.

Turning to adjectives in dual target NPs, all groups preferred using singular adjectives regardless of the noun number, as shown in Table 2.7. The effect of group on singular adjectives in dual NPs did not reach significance (F=2.673; p<.09), however there was a tendency for the HP group to use singular adjectives with dual nouns less than the LP group.

Table 2.7. Mean % adjective number for dual target noun produced, for each group

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
<th>Natives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>100</td>
<td>73</td>
<td>61.11</td>
</tr>
<tr>
<td>Dual</td>
<td>0</td>
<td>26.66</td>
<td>11.11</td>
</tr>
<tr>
<td>Plural</td>
<td>0</td>
<td>0</td>
<td>27.77</td>
</tr>
</tbody>
</table>
To summarize, variability in the production of dual NPs is observed not just in L2ers but also even in native speaker’s production. The preference for singular adjectives, even if the noun has the dual morpheme, was found across all groups.

### 2.2.2 Comprehension Results

The comprehension task produced responses to 240 NPs. The accuracy rates for category and feature by group are presented in Table 2.8 below.

<table>
<thead>
<tr>
<th>Table 2.8. Mean % accuracy on the comprehension task, by proficiency group and feature</th>
<th>LP</th>
<th>HP</th>
<th>Native</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender (all)</td>
<td>90.27</td>
<td>94.31</td>
<td>100</td>
</tr>
<tr>
<td>Masculine</td>
<td>88.88</td>
<td>97.72</td>
<td>100</td>
</tr>
<tr>
<td>Feminine</td>
<td>91.66</td>
<td>90.90</td>
<td>100</td>
</tr>
<tr>
<td>Number (all)</td>
<td>88.88</td>
<td>96.59</td>
<td>100</td>
</tr>
<tr>
<td>Singular</td>
<td>97.22</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>Dual</td>
<td>80.55</td>
<td>93.18</td>
<td>100</td>
</tr>
</tbody>
</table>

Beginning with the effect of proficiency on number, the ANOVA revealed a main effect of proficiency (F=9.13, p<.001). Post hoc tests indicate that the effect of proficiency is attributable to a difference between the LP and HP groups (p<.05) and between the LP and native groups (p<.001). The HP and native groups are at or near ceiling, with only the LP group showing clear evidence of variability in the comprehension of number agreement. The LP group did show numerically higher accuracy rates on singular than dual number, however this difference was not significant (t=0.024; p<.9). Similarly, although the HP group’s accuracy on dual was numerically higher than that of the LP group, the difference was not significant (t=0.92; p<.9).
Turning to gender agreement in comprehension, the results of the ANOVA revealed a main effect of proficiency (F=4.058; p<.02). Post hoc analysis showed this effect to be driven by the difference between native and LP groups (p<.02). The interaction between proficiency and gender was not significant (F=.938; p<.1). The effect of feature gender was not significant (F=0.985; p<.5). In general, it appears that gender was less problematic for L2ers in the comprehension task.

2.3 Discussion

This pilot study aimed to understand what the general features of L2 speakers’ knowledge of gender and number agreement in Arabic NPs, and what error patterns they exhibit. First, consistent production of morphological agreement in the NP appears to be a persistent problem in Arabic L2ers, in line with research on other L2 languages. Even learners at advanced proficiency levels show variability in inflectional morphology. In both tasks there was a gradual decrease in morphological variability in gender agreement with increasing proficiency. The expected advantage for masculine in production was confirmed statistically in the LP group (91.66% accuracy for masculine compared to 50% for feminine), and was qualitatively present though reduced in the HP group. L2ers overuse masculine adjectives with feminine nouns, but do not use feminine adjectives with masculine nouns\(^5\). Below are examples of gender and number errors in NPs made by LP and HP groups:

*Participant 1, Low Proficiency L2 Arabic:*

11. (a) IL: *qitta sagheer*
    cat.Sg.F small.Sg.M

\(^5\) One case from the LP group used a feminine adjective with a masculine noun.
Examples 11 and 14 show that the participants used masculine adjectives instead of feminine ones, which indicates that variability is still persistent at the advanced stages of proficiency. It also shows that masculine inflection occurs in feminine contexts but not the reverse. Examples 12 and 13 show the use of masculine adjectives in feminine contexts and the use of singular number in dual contexts, but the reverse pattern was not attested. The variability observed in the production task thus confirms results from previous L2 research (Bartning, 2000; Bruhn de Garavito, & White, 2002; Franceschina, 2001; Lardiere, 1998; McCarthy, 2008; Santoro, 2012 and White et. al, 2004).

The picture with number accuracy shows some parallels, but also some interesting differences. In contrast to previous studies (Bruhn de Garavito, & White, 2002; Franceschina, 2001; Lardiere, 1998; McCarthy, 2008; Santoro, 2012 and White et. al,
2004), production accuracy on number agreement was significantly lower than gender agreement at all levels. This was clearly driven by problems with the dual. Singular was unproblematic, further supporting the idea that L2er utilize default morphology (here singular, masculine; see McCarthy 2007 for additional discussion of this issue). One potential explanation for the special problem with dual is simply that this feature is not available in English, and thus not facilitated via transfer. However, this does not straightforwardly explain why dual number would be more problematic than gender given that the latter is not present in the L2 either.

The most surprising aspect of our results come from native speakers, who only produce 8.3% of target dual NPs with dual morphology on the adjective (Table 2.4). Across all groups, singular adjective agreement was used for dual targets, even when the dual nominal morphology was correctly produced (Table 2.7). For the L2ers, these results confirm the general overuse of singular morphology found in previous studies. However, for the NS group, there is something more complex going on. Recall that the stimuli in the production task were comprised of animate and inanimate nouns. Table 2.9 and Table 2.10 break down production results for adjectives with dual animate compared to inanimate nouns for all groups.

<table>
<thead>
<tr>
<th>animate</th>
<th>LP</th>
<th>HP</th>
<th>Natives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>100</td>
<td>75</td>
<td>25</td>
</tr>
<tr>
<td>Dual</td>
<td>0</td>
<td>25</td>
<td>12.5</td>
</tr>
<tr>
<td>Plural</td>
<td>0</td>
<td>0</td>
<td>62.5</td>
</tr>
</tbody>
</table>
Table 2.10. Average % production of adjective with each number agreement type
number with dual inanimate nouns by proficiency group

<table>
<thead>
<tr>
<th>inanimate</th>
<th>LP</th>
<th>HP</th>
<th>Natives</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>100</td>
<td>71.42</td>
<td>90</td>
</tr>
<tr>
<td>Dual</td>
<td>0</td>
<td>28.5</td>
<td>10</td>
</tr>
<tr>
<td>Plural</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

For L2 learners, the pattern of results does not depend on animacy—singular adjectives are the prevailing choice in either case. However, native speakers clearly prefer using plural adjectives with dual animate nouns but singular adjectives with dual inanimate nouns. This can be understood in relation to the Arabic plural agreement paradigm described in section 1.2.2 in chapter 1.

When the plural noun refers to a human, the adjective must always agree in number whereas a non-human (animate or inanimate) referent invariably triggers feminine singular. The native speakers in the experiment appear to classify animate dual nouns with human plurals (i.e. trigger plural adjective agreement) but inanimate dual nouns as non-human plurals (i.e. triggering the invariant singular feminine). This suggests the possibility that dual may be in the process of being lost in Arabic. Indeed, Ferguson (1959) states that duals are assimilating to plural in varieties of Arabic other than Cairene Arabic and Modern Standard Arabic. Additional research is needed to investigate the apparent change in the number system the variety of Arabic spoken in Saudi Arabia. In short however, this result does not indicate any problems with our stimuli, but rather a difference between the Modern Standard Arabic taught to our L2 participants, and the native Arabic variety used by our control group.
Turning to the critical question of whether morphological variability is attested in comprehension as well as production, this study confirms the predictions of the MUSH to some extent. For example, the LP group struggle in comprehension relative to the HP and native groups in both gender and number. According to McCarthy (2008) if production and comprehension variability are qualitatively similar, i.e. L2ers use masculine defaults in feminine contexts and singular defaults in plural ones in both tasks, then morphological variability is more than just a performance issue, but a problem with representing the features of gender and number. However the results in this study only partially support McCarthy’s (2008) claim. First, the level of variability in comprehension is markedly lower than that found in production. This decrease across tasks is in line with the idea, put forth by supporters of the MSIH (Prevost & White, 2000) that decreasing communication pressure should decrease variability—this was supported by the results of this study. Second, it is not clear that the pattern of errors is the same across production and comprehension. In the production task, it is clear that L2ers overuse masculine and singular, suggesting these are treated as defaults. On the other hand, in the comprehension task, no clear pattern emerges. Analyzing gender errors in comprehension, L2ers were approximately equally likely to choose the feminine mismatch in masculine target contexts and the masculine mismatch in feminine target contexts, albeit at low rates for each case. This suggests that variability in comprehension exists, but may not in this case be qualitatively similar to production.

One potential factor in the asymmetry between production and comprehension was the stimuli used across both tasks. As described above, the production task used non-
human inanimate objects, while in the comprehension task the pictures were all human. Therefore, the asymmetry in stimuli between production and comprehension could be attributed to this difference. According to Moawad (2006), in L1 acquisition of Arabic, animate human plurals and duals and are acquired before non-human or inanimate plurals and duals. L2ers may be following the same trajectory: they are particular poor with non-human inanimates (at least in production). An alternative explanation is the frequency of the stimuli words. Arabic L2ers receive input from the classroom, textbook and/or instructor. During the production task, if a word is highly frequent in their input, L2ers might provide the correct inflection on nouns and adjectives. If a word is not frequent, L2ers might fail to produce proper inflection on the nouns and adjectives.

2.4 Conclusion

The results of Experiment 1 provide an important starting point for further investigation along two main dimensions. First, the study revealed higher levels of morphological variability in number compared to gender, contrary to many previous studies (Alhawary, 2005; 2009; Bartning, 2000; Bruhn de Garavito, & White, 2002; Franceschina, 2001; McCarthy, 2007; 2008, Montrul, 2008; White et al, 2004). It seems clear that this difference is related to the dual distinction—the vast majority of studies have focused on the acquisition of singular and plural agreement by speakers whose L1 makes those distinctions. Here we have tested L2 acquisition of a number feature by a population of speakers whose L1 does not contain that feature. This result has important implications for relating L2 learning to theories of morphological features. Interestingly, the experiment also revealed that native Arabic speakers use the dual (at least in
production) in a way that likely does not match up with the input to L2 acquisition in a classroom setting.

Second, the particular problem L2ers appear to have with non-human plurals suggests the need for stimuli properties to be identical across production and comprehension tasks. Therefore in the remaining studies, animacy will be explicitly manipulated. Along these same lines, the following experiments alter the procedure described here in a number of ways to hold both the task itself and the knowledge required to complete it as constant as possible. Error rates and error patterns in the two tasks will then be analyzed carefully. Error patterns across both tasks provide insight into the learner’s IL grammar. If the error patterns in both tasks are qualitatively similar, then the source of morphological variability in L2ers is not simply a performance issue, but an underlying competence issue (McCarthy, 2008).

This chapter has revealed thus far the results of Experiment 1; a pilot experiment conducted on gender and number agreement in L2 Arabic NPs. Results of experiment 1 showed that morphological variability is a persisted problem even at advanced proficiency levels. Results also indicate that morphological variability extends to comprehension, however this experiment does not show if they are qualitatively similar or not, therefore cannot support predications of the MUSH (McCarthy, 2007). Results of experiment 1 did support the MSIH (Prevost & White, 2000), reducing communication pressure reduces variability. This can be seen in the higher accuracy rates of comprehension for both proficiency groups. For this, it is the aim of experiment 2 in chapter 3 to further test the predictions of the MUSH (McCarthy, 2007) and the
predictions of the MSIH (Prevost & White, 2000). In addition, experiment 1 followed the methodology by McCarthy (2007;2008) which did not take into consideration the complete Arabic number system or effects of animacy. Therefore, experiment 2 in chapter 3 has identical stimulus across both tasks and animacy is manipulated to test the effects of animacy on the acquisition of feature agreement.
Chapter 3. Morphological Variability in L2 Arabic NPs

This chapter addresses the acquisition of Arabic agreement in the nominal domain, specifically in NPs that consist of a head noun and an attributive adjective. The focus will be on gender (masculine/feminine) and the complete number system (singular/dual/plural) by native English speakers. However, unlike the stimuli and procedure in experiment 1 in chapter 2, here stimulus properties will be identical across production and comprehension tasks, animacy will be explicitly manipulated, and the procedure itself will be held as constant as possible. Recent studies on Morphological variability in second language acquisition have shown that animacy and semantic gender play a role in the acquisition of agreement. Alarcon (2010) found that L2 learners of Spanish are sensitive to linguistic cues such as semantic gender and animacy to establish correct agreement features. In addition, error rates and error patterns in the two tasks will be analyzed carefully in order to provide insight into the learner’s interlanguage grammar. If the error patterns in both tasks are qualitatively similar, then the source of morphological variability in L2ers is not simply a performance issue (cf. MSIH, Prevost & White, 2000), but an underlying competence issue (cf. MUSH, McCarthy, 2007).

The experiment in this chapter aims to address the research questions that were established in chapter 1 (repeated below) in addition to question 4 that seeks to test the role of animacy on acquisition of gender and number agreement in L2 Arabic NPs:
1. What are the general features of L2 speakers’ knowledge of gender and number agreement in Arabic NPs?

2. What are the error patterns of agreement morphology in terms of gender (masculine, feminine) and number (singular, dual) within NPs?

3. Is morphological variability in the production task similar to morphological variability in the comprehension task?

4. What is the role of animacy (human vs. non-human) in the acquisition of Arabic L2 gender and number features?

3.1 Experiment 2

The aim of the study was to provide a quantitative and qualitative comparison of errors in comprehension and production, and to investigate the development of knowledge of the complete Arabic number and gender system in nouns across proficiency levels.

3.1.1 Participants

40 speakers of L2 Arabic were recruited from Arabic language courses at major universities in Northern Virginia and Maryland. All were native speakers of English with no exposure to Arabic prior to enrolling in classes at the university. Following Alhawary (2005, 2009), participants were grouped according to their placement in the Arabic program at their institution, and their length of exposure to Arabic. Group 1 (low proficiency) were all in their 2\textsuperscript{nd} year of the Arabic program, and had received between 135-180 hours of formal instruction in Arabic. Group 2 (high proficiency) were in their 3\textsuperscript{rd} or 4\textsuperscript{th} year of the program and had received more than 225 hours of instruction. A
control group of 5 native speakers of Arabic (excluding heritage speakers) was also tested. Due to the discrepancies between spoken Arabic and instructional Arabic, all members of the control group were teachers of Arabic at the university level. The control group was asked to speak Arabic used in a classroom context and not their specific dialect.⁶

3.1.2 Elicited Production

In the production task, L2 Arabic speakers were shown pictures and were asked to produce noun-adjective combinations. The goal of the task was to examine gender and number agreement in production.

Materials

Materials for the elicited production task include two vocabulary sets. Each set included 48 colored pictures each intended to elicit the production of a noun plus an adjective. Half of the nouns were masculine, half were feminine and within each gender category, the stimuli were divided between presentation as singular, plural, or dual number. To test the effect of animacy on performance, half of the pictures depicted humans and half non-human pictures (split evenly between animate and inanimate). Table 3.1 below provides a sample complete list of test items (for the complete list of test items, refer to Appendix B). Table 3.2 provides a list of adjectives which the pictures were meant to elicit. These adjectives agree with the head noun in gender and number as shown. An addition 20 pictures, intended to elicit verb phrases, were used as filler items.

---

⁶ The control group consisted of Arabic instructors from different countries such as Egypt, Morocco and Palestine.
<table>
<thead>
<tr>
<th></th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Human</strong></td>
<td>Masc mohandis ‘engineer’</td>
<td>Masc mohandisan ‘2 engineers’</td>
<td>Masc mohandisun ‘engineers’</td>
</tr>
<tr>
<td></td>
<td>Fem raqiya ‘dancer’</td>
<td>Fem raqisatan ‘2 dancers’</td>
<td>Fem raqisataat ‘dancers’</td>
</tr>
<tr>
<td></td>
<td>riyadya ‘athlete’</td>
<td>t’aaliba ‘student’</td>
<td>t’aalibatan ‘students’</td>
</tr>
<tr>
<td></td>
<td>tºaaliba ‘student’</td>
<td>tºaalibatan ‘students’</td>
<td></td>
</tr>
<tr>
<td><strong>Non-</strong></td>
<td>qalam ‘pen’</td>
<td>sayyarat ‘2 pens’</td>
<td>aqlam ‘pens’</td>
</tr>
<tr>
<td><strong>human</strong></td>
<td>kalb ‘dog’</td>
<td>qittatan ‘2 dogs’</td>
<td>kilab ‘dogs’</td>
</tr>
<tr>
<td></td>
<td>qittat ‘cat’</td>
<td></td>
<td>qitta ‘cats’</td>
</tr>
</tbody>
</table>

Recall that in Arabic, nouns are marked for gender and number, and adjectives agree with the nouns they modify in these features. However, the agreement paradigm for non-humans differs from that of humans: adjectives modifying non-human plural nouns always take *singular feminine* agreement. In other words the gender and number
distinctions on adjectives are neutralized for non-human plural nouns as in sentence 15 below.

15.  (a) makatib qadeem–a desk.Pl.M old.Sg.F ‘old desks’
     (b) t’awila–at qadeem–a table.Pl.F old. Sg.F ‘old tables’

Procedure

Participants were first asked a set of demographic questions prior to administering the task (for example, “What is your native language?”; “Do you speak any other languages fluently?”; for a full list, refer to Appendix A). To ensure that the type of task (comprehension or production) and lexical items had no effect on the participants’ performance, half the participants completed production first and the other half comprehension first. Within those groups, to ensure that the particular set of lexicon items did not influence the results, participants were divided equally between vocabulary sets 1 and 2.

The task was presented using PsychoPy software (Peirce, 2009), with pictures presented in a randomized order for each participant. On each trial, the target picture was presented along with two distractors which differed in number. This was done to maintain the same procedure across production and comprehension tasks, and to ensure that participants understood that number was relevant to their description. The left picture was always singular, the middle picture was dual and the right picture was plural. The target item was highlighted with a green rectangle. Figure 3.1 below shows a sample of a dual, masculine, human test item.
The main objective of the study was to test participants’ ability to produce agreeing noun + adjective phrases, rather than to test their ability to name lexical items. Therefore, the experimenter provided the noun in its citation form (singular) at the beginning of each trial. Note that this also makes the production task more similar to the comprehension task (in which the agreement features, rather than the noun itself are at issue). Participants were instructed that they had to use that noun to provide an accurate description of the highlighted picture. The experimenter would ask a question to elicit the noun number as in 16. The participant would respond by naming the inflected noun as in 17 below.

16. man fii al-Soora?
    Who in the-picture?
17. haaða ẗ'abeeb
    this a doctor (m).

Immediately after naming the noun, the experimenter asked a question to elicit an adjective as in 18 and participant’s response would be to name the noun plus an adjective as in 19.
18. awSif al-t'abeeb
   describe the-doctor (m)
19. t'abeeb kabeer
   doctor   old

Participants were given 4 familiarization trials before the task began to be sure they understood the procedure. Each set of pictures was displayed for 60 seconds unless the participant clicked the ‘space’ bar to move onto the next trial. In their descriptions of the objects, participants used NPs containing adjectives (describing size, shape, color, depth and length) and the head noun. Each session with each participant was audio recorded using Audacity Version 2.0 (2013), transcribed and coded by the experimenter.

3.1.3 Comprehension

The comprehension task tested L2 Arabic speakers’ comprehension of gender and number of agreement in a picture identification task. This task tests the learners’ ability to identify the correct picture from the number and gender features in a noun phrase.

Materials

Test materials for the comprehension task consisted of 48 target sentences with noun + adjectives phrases. As in the production task, nouns were balanced for gender (masculine, feminine) and number (singular, plural, dual), and half the target test items were human and half were non-human (evenly split between animate and inanimate). For each target phrase, the correct picture was displayed along with two distractors. For human target items these included a picture with the incorrect gender, and a picture with the incorrect number. For non-human target items, grammatical gender was not manipulated because such nouns cannot vary in gender (they are either masculine or
feminine). For these cases, distractor pictures included two additional pictures with the incorrect numbers. An additional 20 sentences consisting of verb phrases were used as filler items.

The noun phrases the participants listened to in the task were embedded in equational (i.e. verbless) sentences consisting of a determiner, a noun and an adjective that show gender (masculine/feminine) and number (singular, dual and plural) agreement. It is important to note that the plural determiner ‘haʔulaaʔ’ (‘these’) does not have gender distinctions, and therefore provides no extra cue to gender for participants. Examples 20, 21 and 22 show the agreement paradigm in equational sentences in singular, dual and plural respectively. These sentences were produced by a native Arabic speaker (the experimenter) and recorded for playback during the experiment.

20. (a) haʔaa rassam maher
this.Sg.M painter.Sg.M talented.Sg.M
“This is a talented painter”
(b) haʔihee momarida – a sayeer-a
this.Sg.F nurse.Sg.F young.Sg.F
“This is a young nurse”
21. (a) haʔaan rassam – an maher - an
“These are two talented painters”
(b) hataan momarida – tan sayeera – tan
this.Dual.F nurse. Dual.F young. Dual.F
“These are two young nurses”
22. (a) haʔulaaʔ rassam – uun maher – uun
these.Pl painter.Pl.M talented.Pl.M
“These are talented painters”
(b) haʔulaaʔ momarida – at sayeera – at
these.Pl nurse.Pl.F young. Pl.F
“These are young nurses”

For singular and dual non-human test items, the agreement paradigm is the same as in sentences 20 and 21 above. Non-human plural nouns (regardless of their gender) are
always preceded by a singular feminine determiner and followed by a singular feminine adjective as in sentence 23 below.

23. (a) haðihee kutub qadeem – a this.Sg.F books.M old.Sg.F “These are old books”
    (b) haðihee sayyara – at sayeer-a this.Sg.F car.Pl.F small.Sg.F “These are small cars”

Procedure

As described above, half the participants completed production first and the other half comprehension first. Within those two groups, participants were again divided equally between vocabulary sets 1 and 2 for the comprehension task.

The task was presented using PsychoPy software (Peirce, 2009), with order of presentation randomized across participants. For each trial, the three pictures were presented and then the stimulus sentence was played. Participants were asked to choose which of three pictures corresponded to the sentence heard by clicking on the appropriate choice. To keep the task as similar as possible to the production task, the picture on the left is always singular and feminine with human nouns only. For non-human cases, the picture on the left was always singular with no gender manipulation. A sample comprehension trial is shown in Figure 3.2 below.
For the trial shown in Figure 3.2, the stimulus was *hađaa mozareʃ* (“This is a farmer”) referring to masculine singular, therefore the correct response is picture number 3.

### 3.2 Results

Accuracy rates were analyzed statistically with a series of mixed effects logistic regression models. Statistical significance for each fixed effect was assessed using likelihood ratio tests, a method of nested model comparison that determines whether model fit is significantly improved by adding additional fixed effects while controlling for added complexity. All analyses were conducting using R (R Core Team, 2013), and the lme4 package (Bates 2010). For each analysis, group is a between-subjects fixed effect with two levels (low, intermediate) gender is a within-subjects fixed effect with two levels (masculine or feminine gender), and number is a within-subject fixed effect with three levels (singular, dual or plural number). Where indicated, analyses may also include a fixed effect of animacy with two levels (human or non-human). All factors were sum coded, meaning individual levels of each factor were compared to the grand mean. Participant, target noun and target adjective are included as random effects for all models.
3.2.1 Production Results

The production task yielded a total number of 1920 tokens of N+Adj phrases. Of these, 1494 tokens were included in the analysis for both gender and number agreement; these were cases in which participants were able to produce the noun with correct gender and number along with an adjective (following McCarthy 2008).

3.2.1.1 Gender Analysis

Figure 3.3 below shows the mean accuracy rates for each group on adjectival gender inflection. Accuracy rates were higher overall for the HP group compared to the LP for both genders. Across both groups, we also find numerically higher accuracy rates for masculine target adjectives compared to feminine. Likelihood ratio tests confirmed a significant effect of proficiency ($\chi^2 (5)=8.79$, p=0.003), but revealed no significant effect of gender ($\chi^2 (6)=1.65$, p=0.19), and no significant effect of the interaction between proficiency and gender ($\chi^2 (7)=0.24$, p=0.62).
As discussed in Chapter 1, Arabic does not maintain gender distinctions on adjectives modifying non-human plurals; all such adjectives take feminine gender. If this is treated by L2 learners as a featural mismatch between masculine plural non-human nouns and feminine adjectives, this combination may lead to particularly low accuracy rates. For masculine non-human nouns, this means a mismatch in gender and number. If this mismatch is problematic, we might expect low accuracy on non-human nouns. Indeed, accuracy rates of adjectival inflection for non-human targets are relatively low for both proficiency groups. However this collapses across target noun number, thus this issue will be revisited after the results on number accuracy are reported.
3.2.1.2 Number Analysis

Figure 3.4 below shows the mean accuracy rates for each group on adjectival number inflection. Across both groups, accuracy rates are dramatically higher for singular target adjectives compared to dual and plural. Likelihood ratio tests assessing the effect of proficiency and target adjective number revealed an overall significant effect for adjective number ($\chi^2(6)=76.21, p<0.001$) and a significant interaction between proficiency and adjective number ($\chi^2(8)=8.33, p=0.015$). Individual comparisons between levels of each factor were conducted using the multcomp package in R (Hothorn et al. 2008). This analysis revealed that for both proficiency groups, plural and dual are significantly worse than singular (LP: pl-sg $\beta=-1.96\pm0.39$, p<0.001; dual-sg $\beta=-3.02\pm0.37$, p<0.001 and for HP: $\beta=-2.63\pm0.33$, p<0.001, pl-sg $\beta=-2.62\pm0.29$, p<0.001). However, dual is significantly worse than plural for the LP group ($\beta=-1.07\pm0.38$, p=0.01) but not for the HP group ($\beta=0.01\pm0.26$, p=0.99). This suggests that the LP group has particular difficulty with dual agreement.
Figure 3.4. Mean % accuracy of adjective inflection on the production task, by proficiency and number across proficiency groups

Again, we expect a decrement in accuracy for non-human plural nouns, where adjective agreement for gender and number are neutralized. If learners treat this as a featural mismatch, we would expect the decrement to specifically target masculine plural non-human target nouns as opposed to feminine plural non-human nouns. This is because in the masculine cases, the adjective mismatches in both gender and number, while in feminine cases the mismatch only occurs on number. This issue is discussed in the following section.
3.2.1.3 Further analysis of human vs. non-human: look into the non-human plural analysis

Figure 3.5 and Figure 3.6 show the mean percent accuracy of adjectival number and gender inflection for singular, dual and plural and human, non-human target nouns for each proficiency group respectively. Recall that when the target noun is non-human masculine or feminine plural the correct adjective takes the feminine singular form. Thus for non-human nouns, there are no cases in which the target adjective agreement is plural. For both proficiency groups, there is a clear decrement in performance for non-human singular feminine target adjectives. This is precisely as expected if mismatching agreement features presents a problem for learners.

Figure 3.5 below shows that LP learners have trouble overall with non-human nouns, but tellingly, accuracy rates on feminine singular adjectival agreement are lower than for masculine singular. Closer analysis reveals that these low accuracy rates are driven by masculine plural non-human target nouns (48%) and feminine non-human plural target nouns (53%). By contrast, when the target noun itself is non-human feminine singular—i.e., when there is no feature mismatch—accuracy rates are much higher (76%).
Figure 3.5. Mean % correct for singular, dual and plural adjective inflection for LP

Likelihood ratio tests for the LP group (Figure 3.5) were used to assess the effect of target noun number, target noun gender and animacy on accuracy. This confirmed a significant main effect of animacy ($\beta = -0.73\pm0.13$, $p<0.0001$), a significant interaction of animacy and target noun gender ($\beta =-0.28\pm0.13$, $p=0.03$) and a significant three-way interaction of animacy, target noun gender and target noun number ($\beta =-0.29\pm0.13$, $p=0.02$). In summary, LP learners have a by lower accuracy rates on non-human noun targets overall, with a particularly low accuracy rate for masculine plurals in the non-human category.

Similar to the LP group, there is a general decrement in performance for non-human nouns in the HP group (Figure 3.6). The low accuracy rates on the singular non-
human feminine target adjectival agreement are again driven largely by masculine non-human plural target nouns (51%). There is less of a problem with feminine non-human plural target nouns (74%), though performance is still much lower than for feminine singular target nouns (90%). Likelihood ratio tests assessing the effect of target noun number, target noun gender and animacy on accuracy in this group confirmed a significant main effect of animacy ($\beta = -0.73\pm0.21$, $p=0.0005$), a significant main effect of target noun number ($\beta = -0.47\pm0.21$, $p=0.02$), a significant interaction of target noun number and target noun gender ($\beta = -0.71\pm0.20$, $p=0.0006$) and a significant interaction of animacy and target noun gender ($\beta = -0.41\pm0.20$, $p=0.04$). These results confirm lower accuracy rates on non-human noun targets overall, and with lower accuracy rates on plural targets overall, though more so with masculine than feminine targets.

Figure 3.6. Mean % correct for singular, dual and plural adjective inflection for HP
To summarize, for both proficiency groups overall accuracy rates for adjective agreement are lower for non-human than human nouns, and non-human masculine plural nouns have the lowest accuracy rates. The results suggest that L2ers have particular difficulty acquiring parts of the agreement paradigm which are neutralized for gender and number distinctions that are otherwise consistently marked on adjectives. They may be treating these cases as essentially involving featural mismatch: plural masculine non-human nouns triggering singular feminine agreement on adjectives. To more clearly understand the mechanism behind L2ers difficulty with these forms, below we look at the type of errors they make in production with respect to non-human plural nouns and dual nouns.

3.2.1.4 Error Analysis

This section analyzes participants’ errors with respect to adjective inflection; in other words, what gender and number they produce, when they get it wrong. Recall that participants are given the noun form, and have to produce an agreeing adjective. As above, here we focus on errors involving incorrect gender and number specification on the adjective when the target noun is non-human masculine plural and the target noun itself is produced with the correct gender and number.

For both groups, participants make more number errors (LP=52%, HP=46%) than gender errors (LP=29%, HP=17%). In line with many studies investigating the potential use of default forms by L2 speakers (e.g., Alacorn, 2010; Martinez-Gibson, 2011 and

7 Note that frequency differences are not a likely explanation for L2ers difficulty with non-human target nouns. These are not low frequency nouns in general, and these learners have been exposed to them in their textbooks for over 1.5 years.
Renaud 2010), both errors involving overuse of the masculine are more common. For human nouns, such gender errors are for the most part driven by the LP group, and tend to occur with plural targets (see Figure 3.5). As noted above, gender errors occur in both groups in the context of non-human targets. Table 3.3 provides a reminder of the adjectival agreement system. Each cell indicates the required adjective form given a noun with the features indicated by column and row headers. For example, a masculine, human, plural noun triggers masculine singular adjectival agreement. On the other hand, a masculine, non-human, plural noun, triggers feminine singular adjectival agreement. Grayed cells indicate forms were gender/number distinctions are neutralized. Importantly, if this is due to problems with featural mismatch, the expectation is that when the target noun is masculine plural, participants may provide a masculine plural adjective, even though a feminine singular is actually required. Indeed, for the LP group, participants switch the adjective to masculine in 24% of these cases and use plural instead of singular in 32%. The HP group behaves similarly (gender switched to masculine: 12% and number switched to singular: 14%). When the target noun is non-human feminine plural, the expectation is that participants may provide the feminine plural, even though the singular is required. This does happen, however the percentage of gender and number errors decreases relative to masculine cases across both groups. Perhaps unsurprisingly, given participants’ difficulty with dual, neither proficiency group exhibit errors involve mistakenly producing dual number adjectives.
Table 3.3. Schematic picture of adjectival agreement paradigm in Modern Standard Arabic

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th></th>
<th>Feminine</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
<td>Non-human</td>
<td>Human</td>
<td>Non-human</td>
</tr>
<tr>
<td>Singular</td>
<td>M.Sg.</td>
<td>M.Sg.</td>
<td>F.Sg</td>
<td>F.Sg</td>
</tr>
<tr>
<td>Dual</td>
<td>M.Dual</td>
<td>M.Dual</td>
<td>F.Dual</td>
<td>F.Dual</td>
</tr>
<tr>
<td>Plural</td>
<td>M.Pl</td>
<td>F.Sg</td>
<td>F.Pl.</td>
<td>F.Sg</td>
</tr>
</tbody>
</table>

We now turn to errors involving incorrect gender and number specification on the adjective when the target noun is dual (again only looking at cases where the target noun itself was produced with the correct gender and number). Table 3.4 and Table 3.5 show adjective number and gender response for both proficiency groups when the target noun is dual masculine and dual feminine respectively.

Table 3.4. Adjective Gender and Number Response when Target Noun Is Dual Masculine

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>29% (default)</td>
<td>14%</td>
</tr>
<tr>
<td>Dual</td>
<td>55% (correct)</td>
<td>0%</td>
</tr>
<tr>
<td>Plural</td>
<td>1%</td>
<td>&lt;1%</td>
</tr>
</tbody>
</table>

Table 3.5. Adjective Gender and Number Response when Target Noun Is Dual Feminine

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>0% (default both)</td>
<td>33% (default number)</td>
</tr>
<tr>
<td>Dual</td>
<td>8% (default gender)</td>
<td>55% (correct)</td>
</tr>
<tr>
<td>Plural</td>
<td>0%</td>
<td>5%</td>
</tr>
</tbody>
</table>

From the tables above it is clear that participants in both proficiency groups are more likely to provide singular adjectives when dual adjectives are required. They rarely
use plural adjectives in place of duals. This is to some extent unexpected since in English
dual is grouped grammatically with plural in English. If learners were treating dual as
plural, they should make errors in which they use plural adjectives with dual nouns.
Instead it appears that dual errors largely reflect use of the default singular feature. That
said, there actually some evidence for the dual being treated as plural. When the target is
masculine, there are a fair number of errors involving production of the feminine
singular. A closer look reveals that these errors are almost exclusively made on non-
human nouns (13% of errors for non-human as compared to 6% for human nouns). This
is the error that would be expected if some learners treat these as plural and produce the
neutralized feminine singular form. A similar pattern is found for feminine target nouns;
errors largely reflect use of defaults, but number errors in which the singular is used are
overinflated for non-human nouns, potentially reflecting the neutralized form that would
be used if these were plurals (42% of errors for non-human as compared to 23% for
human nouns).

3.2.1.5 Discussion

The results from the production task indicate that morphological variability
remains a persistent problem for L2 learners even at advanced stages of proficiency
(Franceschina, 2001; Lardiere, 1998 and White, 2003). To summarize, L2ers were better
with gender than with number agreement with N+ADJ agreement.

The following are selected examples of gender errors in adjectives (24-28) made
by both LP and HP groups. The examples include a dialogue between the experimenter
(E) and the participant (P). Note again that knowledge of gender of the target noun plays
no role (as in McCarthy 2007) since participants were given the noun. All examples below are human singular masculine and feminine targets nouns.

*Participant LP3, Low L2 Arabic*

24. E: awSif altˤaalib-a  
   describe the-student-F  
   P: *altˤaaliba mabsoot  
   student.Sg.F happy.Sg.M

*Participant HP4, Advanced L2 Arabic*

25. E: awSif alraqiS-a  
   describe the-dancer-F  
   P: *alraqiS-a taweel  
   dancer.Sg.F tall.Sg.M

*Participant LP8, Low L2 Arabic*

26. E: awSif alustað  
   describe the-teacher.M  
   P: *alustað kabeer-a  
   the-teacher.Sg.M old.Sg.F

*Participant LP17, Low L2 Arabic*

27. E: awSif almohandes  
   describe the-engineer.M  
   P: almohandes saʕeer-a  
   the-engineer.M young.F

*Participant HP12, Advanced L2 Arabic*

28. E: awSif alriyady  
   describe the-athlete.M  
   P: alriyady taʕbaan-a
the-athlete.Sg.M  tired.Sg.F

Examples 24 and 25 reflect a relatively common error type: the participants are using masculine adjectives in feminine contexts. Examples 26 through 28 reveal the occurrence of feminine adjectives in masculine contexts. Such errors are to some degree unexpected, given that most studies of L2 morphological learning reveal a strong effect of the masculine default. This effect is present, but not as strong for Arabic L2ers. Interestingly, the emergence of feminine rather than masculine morphology as the default has been attested in Arabic heritage speakers. Benmamoun, Albrini, Montrul and Saadah (2014) report that heritage speakers of Arabic deploy the feminine plural suffix as their default in all agreement patterns. Similar findings have been reported in a number of studies in Arab monolinguals (Albrini, 2015; Omar, 1973 and Ravid & Farah, 1999). These occurrences fail to support McCarthy’s (2007) Morphological Feature Underspecification Hypothesis (MUSH) where she states that L2 morphological errors are not random, but systematic. McCarthy’s (2007) MUSH also states that when L2ers make gender or number errors they will select the underspecified form i.e. masculine, singular. Although the L2 data show evidence for a masculine default, the fact that it is weaker than expected, along with other evidence for the role of feminine as a possible default, call into question whether masculine is always the default form. I will return to this in section 3.3 after discussing the comprehension results. Sentences 29 and 30 below are additional selected examples of non-human masculine singular target nouns produced with feminine adjectives.
29. Participant LP6, Low L2 Arabic

E: awSif alqalam
describe the-pen.M
P: *alqalam jadeed-a
     the-pen.Sg.M new.Sg.F

30. Participant HP7, Advanced L2 Arabic

E: awSif alhisan
describe the-horse.M
P: *alhisan kabeer-a
     the-horse.Sg.M big.Sg.F

Number variability is also found across all L2ers, including advanced speakers. In all singular contexts, errors were rare (less than 5%). However errors with plural and dual forms were well-attested. The relatively poor performance on plural across both proficiency groups runs counter to previous studies (Alhawary, 2005; 2009; Bartning, 2000; Bruhn de Garavito, & White, 2002; Franceschina, 2001; McCarthy, 2007; 2008, Montrul, 2008; White et al, 2004). However, in this case it is clear that the problem stems from a particular set of nouns for which there is neutralization in the agreement paradigm (see Table 3.3). When the results were broken down by gender, number, and animacy, plural accuracy rates for singular and plural human target nouns were consistently high. They were lower over all for non-humans and particular for non-human masculine plural nouns. This suggests that L2ers do not have a problem with plural in general, but rather with the agreement pattern of the non-human masculine plural nouns. The errors
participants made further suggest that the difficulty lies in the featural mismatch involved—these nouns trigger feminine singular agreement.

**3.2.2 Comprehension Results**

As for the comprehension results, comprehension accuracy rates were analyzed statistically with a series of mixed effects logistic regression models and likelihood ratio tests. For each analysis, group is a between-subjects fixed effect with two levels (low, intermediate) and feature values are within-subjects fixed effects with either two (masculine or feminine gender) or three (singular, dual or plural number). Some models also include a fixed effect of animacy with two levels (human or non-human). For comprehension results, since the participant is only listening to a NP and choosing the correct corresponding picture, accuracy rates were only based on target nouns only.

**3.2.2.1 Gender Analysis**

The analysis of gender only included the human tokens since gender is an arbitrary feature of non-human nouns and thus cannot be manipulated. A total number of 1920 tokens were analyzed. Figure 3.7 below shows the mean accuracy by proficiency group and gender. Likelihood ratio tests assessing the effect of noun gender and proficiency on accuracy revealed a significant main effect of proficiency ($\chi^2(4)=6.16$, p=0.01), no main effect of gender ($\chi^2(5)=0.071$, p=0.79), and a significant interaction between proficiency and gender ($\chi^2(6)=4.046$, p=0.044). These results indicate that overall, accuracy rates were higher for the HP group than the LP group, and that the effect over gender changes across these two groups. Looking at Figure 3.7, it appears
that the LP have a slightly higher accuracy level for feminine targets while this reversed for the HP group.

![Mean % Accuracy during the comprehension task, by proficiency and gender.](image)

**Figure 3.7. Mean % Accuracy during the comprehension task, by proficiency and gender.**

### 3.2.2.2. Number Analysis

For the analysis of number, 1920 total tokens were analyzed. Figure 3.8 below shows mean accuracy rates by proficiency group and number. Likelihood ratio tests were used to assess the effect of proficiency and noun number on accuracy. These revealed a significant main of both proficiency ($\chi^2(4)=26.44$, $p<0.0001$) and number ($\chi^2(6)=13.565$, $p=0.001$), but no significant interaction ($\chi^2(8)=0.944$, $p=0.62$). Individual comparisons across levels of each factor revealed that for both proficiency groups, plural is significantly worse than singular only (LP: $pl-sg \beta = -1.13 \pm 0.35$, $p=0.002$; and for HP: $\beta = -1.135\pm0.34$, $p=0.003$).
These results indicate that overall, the HP group had higher accuracy rates than the LP group. Interestingly, in both groups, dual targets have numerically higher accuracy rates than plurals targets. In part, low accuracy rates for the plural may again be attributed to non-human plural target nouns, as in the production task. We will investigate this in the following section. In any case, however, the fact that plural targets had lower accuracy than dual is surprising—and it differs clearly from the production task, where dual was particularly problematic for the LP group.

3.2.2.3 Further Analysis of Human vs. Non-Human

Plural number in comprehension is dramatically lower than singular and dual, again driven by the featural mismatch between the noun and the adjective for non-human nouns. Figure 3.9 and Figure 3.10 below show the mean accuracy rates of noun inflection.
by the feature human vs. non-human for each group. For both proficiency groups, plural number for non-human nouns is the most problematic.

**Figure 3.9. Mean % accuracy on noun inflection by number and animacy for LP**

For the LP group, there is a general effect of animacy on noun inflection. Overall, the LP group has lower accuracy rates on non-human noun targets than on human noun targets with the worst category being non-human plural. A likelihood ratio test assessing the effect of target noun number and animacy on accuracy in the LP group confirms a significant effect of animacy ($\beta = -0.84 \pm 0.106, p<0.001$), dual number ($\beta = 0.49 \pm 0.155, p=0.001$), plural ($\beta = -0.48 \pm 0.14, p<0.001$) and a significant interaction between animacy and plural number ($\beta = -0.31 \pm 0.14, p=0.03$). The latter highlights again the special problem L2ers have with non-human plurals.
A parallel analysis on the HP group data also indicates a general effect of animacy ($\beta = -0.71\pm0.13$, $p<0.0001$), dual number ($\beta = 0.66\pm0.21$, $p=0.001$), plural number ($\beta = -0.71\pm0.16$, $p<0.0001$), but no significant interaction between animacy and target noun number.

It is clear that for both proficiency groups non-human plural nouns are the most problematic—just as in the production task. These results confirm that L2ers are in fact treating the cases of non-human plural as featural mismatches between the noun and adjective. This leads to trouble identifying the correct picture in the comprehension task, and trouble producing the right adjectival agreement in the production task. As we will discuss below, this pattern thus does not appear to be due to production-related pressures but to an underlying representational issue that affects comprehension as well.
3.2.2.4 Error Analysis

This section analyzes participants’ errors with respect to number in the comprehension task. Gender will not be analyzed since accuracy rates were uniformly very high in the comprehension task. Table 3.6 below shows errors in which participants chose the singular or plural distractor for a dual noun target. For both proficiency groups, participants were much more likely to choose the plural rather than singular distractor picture. This pattern of errors is in contrast to what we saw in the production task. It could potentially reflect the fact that the dual suffix is not fully acquired, so when participants hear any suffix on the noun (dual or plural), they interpret (or mistake) this for the plural since the singular form is bare in Arabic.

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>23%</td>
<td>18%</td>
</tr>
<tr>
<td>Plural</td>
<td>77%</td>
<td>82%</td>
</tr>
</tbody>
</table>

For plural nouns, as shown in Table 3.7 below, both proficiency groups use singular nouns in place of plural nouns. On the face of it, these results are somewhat surprising: we might expect that errors would involve choosing dual, since dual and plural are collapsed in English, and as noted above, plural and dual are similar in Arabic in both involving overt morphological marking. However, it is clear that the poor accuracy rate in the plural is driven by non-humans; participants choose singular number rather than dual for non-human plural targets, as shown in Table 3.8.
Table 3.7. Number errors for plural nouns by proficiency

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>74%</td>
<td>76%</td>
</tr>
<tr>
<td>Plural</td>
<td>26%</td>
<td>24%</td>
</tr>
</tbody>
</table>

In the context of the comprehension task, we can understand this by recalling that participants listen to a sentence containing a noun + adjective and they must choose the picture that corresponds to the sentence. For all singular and dual targets, the noun and adjective are both always inflected for corresponding gender and number distinctions. On the other hand, non-human plural nouns are followed by singular feminine adjectives. If participants are using the adjective morphology to determine their choice, they will incorrectly choose the picture with singular number.

Table 3.8. Number substitution for non-human plural nouns by proficiency

<table>
<thead>
<tr>
<th></th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>80%</td>
<td>78%</td>
</tr>
<tr>
<td>Dual</td>
<td>20%</td>
<td>22%</td>
</tr>
</tbody>
</table>

3.2.2.5 Discussion

The comprehension task, like the production task, produced a similar pattern; the L2 learners across both proficiency groups were better able to identify gender better than number agreement within NPs. This trend persisted in both human and non-human nouns. In addition, although not statistically significant, L2 learners’ accuracy rates were
numerically higher for masculine than feminine agreement. When analyzing errors in the feature gender, feminine gender emerges in masculine contexts. These contexts were also human masculine singular target nouns to control for animacy effects.

While singular number was unproblematic across proficiency levels, dual and plural number were clearly problematic for learners, with accuracy on plural number being particularly low for the LP group as opposed to the production task. The relatively poor performance on plural across both proficiency groups stems from a particular set of nouns for which there is neutralization in the agreement paradigm similar to what has been observed in the production task. When the results were broken down by gender, number, and animacy, plural accuracy rates for singular and plural human target nouns were consistently high. They were lower for non-human masculine plural nouns. This suggests that L2 learners do not struggle with plural in general, but rather with the agreement pattern of the non-human masculine plural nouns.

In regards to dual, L2 learners were more successful in the comprehension task than the production task. This greater accuracy supports the predictions of the MSIH (Prevost & White, 2000) that reducing communication pressure will reduce variability.

3.3 General Discussion

To answer the first research question: *What are the general features of L2 speakers’ knowledge of gender and number agreement in Arabic NPs?* Looking at accuracy rates for both production and comprehension tasks with respect to gender reveals that gender features are present in these L2 speakers’ IL. Variability in gender agreement decreases with increasing proficiency in both tasks, but remains present even
at the advanced level. Accuracy rates on number features, on the other hand, are significantly lower than gender for these L2 speakers. While singular number is unproblematic across all proficiency levels, dual and plural numbers were clearly problematic for learners. The higher accuracy rates for gender and poorer accuracy rates for number runs counter to many previous studies (Alhawary, 2005; 2009; Bartning, 2000; Bruhn de Garavito, & White, 2002; Franceschina, 2001; McCarthy, 2007; 2008, Montrul, 2008; White et al, 2004). However, poor accuracy rates in plural can be largely attributed to the non-human masculine subclass of nouns, pointing to the importance of studying cases of neutralization in the agreement paradigm. In addition to gender and number, results across both tasks also indicate an overall effect of animacy, which we will now discuss in further detail.

Results in both tasks confirm that L2 learners are sensitive to animacy when acquiring gender and number features, specifically agreement features between a head noun and attributive adjectives. This is clearly visible when the target nouns are human. Accuracy rates are higher than non-human nouns. Alkhohlani (2016) states the absence of morphological and semantic gender cues on the noun, in this case non-human nouns, causes ambiguity for these L2ers. In these cases, learners tend to overgeneralize the feminine gender to both masculine and feminine nouns. Previous studies support these findings and the effect of animacy; according to Alarcon (2010), L2 speakers are influenced by linguistic cues such as animacy when acquiring grammatical gender. Alarcon (2010) found that L2 speakers use these linguistic cues for establishing correct gender agreement, and that Spanish L2 speakers are sensitive not just to linguistic cues,
but also to overtness. She found that the nouns that were overtly marked –o for masculine and –a for feminine were the ones that were assigned correct gender and agreement. She claims, “animacy and overtness work together to reinforce correct gender agreement” (Alarcon, 2010, p.285). Alarcon’s findings can help explain why L2 speakers in this experiment have low accuracy rates especially when the target noun is non-human masculine plural. One reason is that these nouns are not derived concatenatively - attaching a plural morpheme overtly to the stem. Their derivation is non-concatenative - mapping a root onto a plural template (McCarthy & Prince, 1990) what is known as the broken plural in Arabic. These derivations do not have morphological transparent markings on the stem in the form of prefixes or suffixes. As in Spanish, then, the lack of transparent markings on non-human plural nouns may lead to less reinforcement of agreement features. In the case of experiment 2, it seems that 2 factors are playing a role in the difficulty of acquiring agreement features between non human plural nouns and the attributive adjective: (1) absence of semantic gender on non human plural nouns and (2) absence of overt plural markings on the nouns.

3.4 Conclusion

The results from experiment 2 in this chapter show that morphological variability in L2 Arabic remains a persistent problem even at advanced levels of proficiency, extending to comprehension. Gender agreement acquisition across both tasks was less problematic than number contrary to many previous studies (Alhawary, 2005; 2009; Bartning, 2000; Bruhn de Garavito, & White, 2002; Franceschina, 2001; McCarthy, 2007; 2008, Montrul, 2008; White et al, 2004). Learners had trouble with dual forms,
particular at lower proficiency levels, though this was attenuated in comprehension (supporting the MSIH). In addition, errors involving use of feminine in masculine contexts occurred in both tasks, but it was not a general pattern. More data needs to be analyzed to see if this pattern occurs elsewhere such as DO clitics — this will be revisited in chapter 4. Finally, the effect of animacy (human vs. non-human) was present in both tasks; producing and interpreting correct gender and number was overall more difficult for non-human targets. Neutralized agreement likewise presented a persistent and substantial problem for learners across both tasks. In the following chapter, we will tests whether these patterns are restricted to adjectival agreement, or also emerge with direct object (DO) clitic agreement in L2 Arabic. This will allow us to strengthen any conclusions regarding similarities and differences in morphological variability across comprehension and production.
Chapter 4. Morphological Variability in L2 Arabic Direct Object Clitic Agreement

This chapter addresses the acquisition of gender and number agreement in direct object (DO) clitics in L2 Arabic. As previously discussed in Chapter 3, Experiment 2, morphological variability remains a persistent problem on adjectival inflection even at advanced levels of proficiency. Experiment 2 results have also revealed that gender agreement is less problematic than number agreement in L2 Arabic, and that animacy has an effect on the acquisition of agreement features between head noun and attributive adjective. Therefore, the aim of this chapter is to test if similar patterns emerge in the acquisition of Arabic L2 DO clitic agreement as well. Previous work on DO clitics in L2 Spanish and L2 French have reported that clitic agreement morphology is fully acquirable despite its absence in the learner’s L1, but that errors remain throughout the acquisition process due to “mapping problems” (Renaud, 2010 and Santoro, 2011). As in Experiment 2, Experiment 3 reported in this chapter analyzes error rates and error patterns in both production and comprehension of DO clitics in order to provide insight into the learner’s interlanguage grammar. If the error patterns in both tasks are qualitatively similar, then the source of morphological variability in L2ers is not simply a performance issue (cf. MSIH, Prevost & White, 2000), but an underlying competence issue (cf. MUSH, McCarthy, 2007).
This chapter is organized as follows: section 4.1 describes the features of DO clitics in Modern Standard Arabic, sections 4.2, 4.3 and 4.4 reports on Experiment 3 and section 4.5 concludes the chapter.

4.1 DO Clitics in Modern Standard Arabic

Direct Object (DO) pronouns in Arabic are realized as clitics, written as suffixes attached to the end of verbs. Like all pronouns in Arabic, DO clitics are marked for person, gender and number. For purposes of this dissertation, only 3rd person is used. Table 4.1 below shows the 3rd person DO clitics in Modern Standard Arabic.

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>-hu</td>
<td>-ha</td>
</tr>
<tr>
<td>Non Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Dual</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>-huma</td>
<td>-huma</td>
</tr>
<tr>
<td>Non Human</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Plural</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Human</td>
<td>-hum</td>
<td>-hun</td>
</tr>
<tr>
<td>Non Human</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

4.1.1 Direct Object Clitic Agreement in Modern Standard Arabic

DOs in Arabic involve agreement in gender (masculine or feminine), number (singular, dual or plural), and person\(^8\) (nominative, accusative, or genitive) with the noun it is referring to. As with noun+adjective constructions, plural agreement in DO clitics exhibits an animacy-based distinction between human and non-human entities. Human

---

\(^8\) In the experiments, all forms are in 3rd person
plurals trigger matching agreement. However, non-human plurals invariably take singular feminine clitics regardless of the object’s gender. Examples 30 – 32 below illustrate agreement, and the distinction between human and non-human plurals.

30. **Masculine Singular and Dual DO Clitics (Human/Non-Human)**
   (a) al-walad  d'araba   at'aalib  
       the-boy  hit      the student.Sg.M

   al-walad  d'arabu-hu
   the-boy  hit – CL.Sg.M

   (b) al-walad  d'araba  al-kalb
   the-boy  hit      the dog.Sg.M

   al-walad  d'arabu-hu
   the-boy  hit – CL.Sg.M

   (c) al-walad  d'araba  at'aalibayn
   the-boy  hit      the student.Dual.M

   al-walad  d'arabu-huma
   the-boy  hit – CL.Dual.M

   (d) al-walad  d'araba  al-kalbayn
   the-boy  hit      the dog.Dual.M

   al-walad  d'arabu-huma
   the-boy  hit – CL.Dual.M

31. **Feminine Singular and Dual DO Clitics (Human/Non-Human)**
   (a) al-walad  d'araba  at'aaliba
   the-boy  hit      the student.Sg.F

   al-walad  d'araba-ha
   the-boy  hit – CL.Sg.F
32. Plural DO Clitics (human/non-human distinction)
(a) al-walad d'araba at'ullah
the-boy hit the student.Pl.M

al-walad d'arabu-hum
the-boy hit – CL.Pl.M

(b) al-walad d'araba at'aalibaat
the-boy hit the student.Pl.F

al-walad d'araba-hun
the-boy hit – CL.Pl.F

(c) al-walad d'araba al-kilaab
the-boy hit the dog.Pl.M
4.2 Experiment 3

The aim of Experiment 3 was to provide a quantitative and qualitative comparison of errors in comprehension and production tasks, and to investigate the development of knowledge of the complete gender and number systems across proficiency levels in L2 Arabic DO clitics.

4.2.1 Participants

40 speakers of L2 Arabic were recruited from Arabic language courses at a major university in Northern Virginia and Maryland. All were native speakers of English with no exposure to Arabic prior to enrolling in classes at the university. Following Alhawary (2005, 2009), participants were grouped according to their placement in the Arabic program at this institution, and their length of exposure to Arabic. Group 1 (low proficiency) were all in their 2nd year of the Arabic program, and had received between 135-180 hours of formal instruction in Arabic. Group 2 (high proficiency) were in their 3rd or 4th year of the program and had received more than 225 hours of instruction.

4.2.2 Elicited Production

In the first task, L2 Arabic speakers listened to a prerecorded sentence and saw a corresponding picture on the screen. They were asked to produce a clitic that
corresponded to the shown picture. The goal of the task was to examine gender and number DO clitic production.

**Materials**

Materials for the elicited production task include two vocabulary sets. Each set included 48 colored pictures each intended to elicit the production of a direct object clitic. Each picture included an agent (a male or female divided equally between trials) carrying out action on a target noun. Half of the target nouns were masculine, half were feminine and within each gender category, the stimuli were divided between presentation as singular, plural, or dual number. To test the effect of animacy on performance, half of the pictures depicted humans and half non-human pictures (split evenly between animate and inanimate). From the non-human category, half the pictures were animate. Table 4.2 provides a complete list of Arabic direct object clitics. An additional 20 pictures, intended to elicit prepositional phrases, were used as filler items.

<table>
<thead>
<tr>
<th>Table 4.2 Direct Object Clitics in Modern Standard Arabic</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Singular</strong></td>
</tr>
<tr>
<td>Human</td>
</tr>
<tr>
<td>Non-human</td>
</tr>
</tbody>
</table>

Recall that in Arabic, just like nouns and adjectives are marked for gender and number, so too are direct object clitics. However, for dual nouns, there are no gender distinctions...
as all nouns have the same clitic –huma. In nonhuman plurals, gender and number are both neutralized and have the clitic –ha.

Procedure

As in Experiment 2, participants were first asked a set of demographic questions prior to administering the task (for a full list, refer to Appendix A). To ensure that the type of task (comprehension or production) and lexical items had no effect on the participants’ performance, half the participants completed production first and the other half comprehension first. Similarly, within those two groups, participants were divided equally between vocabulary sets 1 and 2 (for the full list of test items, refer to Appendix C). The task was presented using PsychoPy software (Peirce, 2009) with pictures presented in a randomized order for each participant. On each trial, the target picture was presented with a prerecorded sentence as shown in Figure 4.1 below. The prerecorded sentence was “There is a boy. The boy is pulling the dog. The boy is pulling ……” After the verb, there is a silent pause during which the participant is to produce the direct object clitic.
The main objective of the study was to test participants’ ability to produce clitics rather than to test their ability to name lexical items and verbs. Therefore, the experimenter provided the prerecorded sentence at the beginning of each trial. Participants were given 4 familiarization trials before the task began to be sure they understood the procedure. Each set of pictures was displayed for 60 seconds unless the participant clicked the ‘space’ bar to move onto the next trial. Each session with each participant was audio recorded using Audacity and transcribed and coded by the experimenter.

4.2.3 Comprehension

The comprehension task tested L2 Arabic speakers’ comprehension of gender and number agreement in a picture identification task. This task tests the learners’ ability to identify the correct picture from the number and gender features on the DO clitic.
Materials

Test materials for the comprehension task consisted of 120 target sentences with verb phrases that contained DO clitics. As in the production task, nouns were balanced for gender (masculine, feminine) and number (singular, plural, dual), and half the target test items were human and half were non-human (evenly split between animate and inanimate). For each target phrase, the correct picture was displayed along with one distractor. For human target items distractors were divided evenly between incorrect gender, incorrect number or both. In the dual cases, only incorrect number was used as a distractor since the clitic is the same for both masculine and feminine. For non-human target items, grammatical gender was manipulated by showing completely different items in the distractor pictures because such nouns cannot vary in gender (they are either masculine or feminine). Similar to the human target nouns, dual target direct object clitics, do not distinguish between gender; therefore, only number was manipulated. An additional 20 sentences consisting of prepositional phrases were used as filler items.

Procedure

As described above, half the participants completed production first and the other half comprehension first. Within those two groups, participants were again divided equally between vocabulary sets 1 and 2 for the comprehension task (See Appendix C for full list).

The task was presented using PsychoPy software (Peirce, 2009), with order of presentation randomized across participants. For each trial, the two pictures were presented and then the stimulus sentence was played. The stimulus sentence contained a
verb phrase and a direct object clitic. Participants were asked to choose which of two pictures corresponded to the sentence heard by clicking on the appropriate choice. A sample comprehension trial is shown in Figure 4.2.

For the trial shown in Figure 4.2 Howa yohebo-hu (“He Loves him”) refers to masculine singular; therefore, the correct response is picture number 1.

4.3 Results

Accuracy rates were analyzed statistically with a series of mixed effects logistic regression models. Statistical significance for each fixed effect was assessed using likelihood ratio tests, a method of nested model comparison that determines whether model fit is significantly improved by adding additional fixed effects while controlling for added complexity. All analyses were conducting using R (R Core Team, 2013), and the lme4 package (Bates 2010). For each analysis, group is a between-subjects fixed
effect with two levels (low, intermediate) gender is a within-subjects fixed effect with two levels (masculine or feminine gender), and number is a within-subject fixed effect with three levels (singular, dual or plural number). Where indicated, analyses may also include a fixed effect of animacy with two levels (human or non-human). All factors were sum coded, meaning individual levels of each factor were compared to the grand mean. Participant, target noun and target clitic are included as random effects for all models.

4.3.1 Production Results

The production task yielded a total number of 1920 tokens of direct object clitics. All tokens were included in the analysis for both gender and number agreement.

4.3.1.1 Gender Analysis Figure 4.3 Figure 4.3 below shows the mean accuracy rates of clitic production for each group. For both groups, accuracy rates are higher for masculine target clitics than feminine target clitics. However, the LP scored slightly higher accuracy rates on masculine target clitics than the HP groups. A mixed-effects model with proficiency and target clitic gender as fixed effects and a likelihood ratio test revealed no significant effect of proficiency ($\chi^2 (4)=1.12, p=0.29$). The model confirmed a significant effect for gender ($\chi^2 (5)=17.13, p<0.001$) and revealed a marginally significant interaction between proficiency and gender ($\chi^2 (6)=3.16, p=0.07$).
4.3.1.2 Number Analysis

Figure 4.4 below shows the mean accuracy rates for each group on clitic production. Across both groups, accuracy rates are dramatically higher for singular and plural target clitics compared to dual. A mixed-effects model with proficiency and target clitic number as fixed effects and a likelihood ratio test revealed an overall significant effect for proficiency ($\chi^2(4)=41.60$, $p<0.001$), and a significant effect for target clitic number ($\chi^2(6)=7.32$, $p=0.02$) and a significant interaction between target clitic number and proficiency ($\chi^2(8)=7.29$, $p=0.02$).
A multiple comparison model revealed that for LP, plural and dual are significantly worse than singular (dual-sg $\beta = -2.30 \pm 0.47$, $p<0.001$; dual-pl $\beta = -1.95 \pm 0.49$, $p<0.001$), but the difference between singular and plural is not significant ($\beta = -0.35 \pm 0.42$, $p=0.68$). For the HP group, only dual is marginally worse than plural (dual-pl $\beta = -3.04 \pm 1.22$, $p=0.03$).

Because, as with adjectives, direct object clitics modifying non-human plural nouns take feminine singular clitics (-ha), L2 learners could treat this as a featural mismatch. This means that we need to analyze plural forms separately based on animacy, rather than confound these two different parts of the system. Table 4.3 gives a schematic picture of the DO clitic agreement paradigm in Modern Standard Arabic. Each cell indicates the required DO clitic form for a given noun with the features indicated by
column and row headers. For example, a masculine, human, plural noun triggers masculine singular DO clitic agreement. On the other hand, a masculine, non-human, plural noun, triggers feminine singular DO clitic agreement. Grayed cells indicate forms were gender/number distinctions are neutralized. We will analyze these potential cases of featural mismatch in the section 4.3.1.3.

### Table 4.3 Schematic picture of DO clitic agreement paradigm in Modern Standard Arabic

<table>
<thead>
<tr>
<th></th>
<th>Masculine</th>
<th>Human</th>
<th>Feminine</th>
<th>Non-human</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Human</td>
<td>Non-human</td>
<td>Human</td>
<td>Non-human</td>
</tr>
<tr>
<td>Singular</td>
<td>M.Sg.</td>
<td>M.Sg.</td>
<td>F.Sg</td>
<td>F.Sg</td>
</tr>
<tr>
<td>Dual</td>
<td>Dual</td>
<td>Dual</td>
<td>Dual</td>
<td>Dual</td>
</tr>
<tr>
<td>Plural</td>
<td>M.Pl</td>
<td>F.Sg</td>
<td>F.Pl.</td>
<td>F.Sg</td>
</tr>
</tbody>
</table>

#### 4.3.1.3 Further analysis of human vs. non-human: a look into the non-human plural analysis

Recall that in Experiment 2, there was a general effect of human vs. non-human on adjectival inflection. A similar pattern is observed in the clitic domain as well, however it is not as robust as in the adjectival domain. **Figure 4.5** and **Figure 4.6** show the mean percent accuracy of clitic production for singular, dual and plural and human, non-human target nouns for each proficiency group respectively.
Likelihood ratio tests for the LP group (Figure 4.5) were used to assess the effect of target noun number, target noun gender and animacy on accuracy. This confirmed a significant main effect of animacy ($\beta = -0.64\pm0.12$, $p<0.0001$), a significant interaction of animacy and target noun gender ($\beta = -0.72\pm0.12$, $p<0.0001$) and a significant three-way interaction of animacy, target noun gender and target noun number ($\beta = -0.36\pm0.12$, $p=0.002$). The low accuracy rates on the singular non-human feminine target clitic agreement are again driven largely by masculine non-human plural target nouns (12%). There is less of a problem with feminine non-human plural target nouns (52%). In summary, LP learners have a lower accuracy rates on non-human noun targets overall,
with a particularly low accuracy rate for masculine plurals in the non-human category.

The issue with low dual accuracy rates is discussed in the error analysis section.

![Bar chart](image)

**Figure 4.6. Mean % correct for singular, dual and plural clitic production for HP**

Similar to the LP group, there is a general decrement in performance for non-human nouns in the HP group (Figure 4.6) the low accuracy rates on the singular non-human feminine target adjectival agreement are again driven largely by masculine non-human plural target nouns (35%). There is less of a problem with feminine non-human plural target nouns (60%). Likelihood ratio tests assessing the effect of target noun number, target noun gender and animacy on accuracy in this group confirmed a significant main effect of animacy ($\beta = -0.24 \pm 0.089$, $p<0.0001$), a significant main effect
of target noun number ($\beta = -0.25 \pm 0.088$, $p < 0.0001$), a significant interaction of animacy and target noun gender ($\beta = -0.42 \pm 0.088$, $p < 0.0001$), but no significant interaction of target noun number and target noun gender ($p = 0.5$). These results confirm lower accuracy rates on non-human noun targets overall, and with lower accuracy rates on plural targets overall, though more so with masculine than feminine targets. Similar to the observation in the LP group, the low accuracy on dual targets is discussed in the following section when exploring the types of errors these L2ers make.

To summarize, for both proficiency groups overall accuracy rates for clitic production are lower for non-human than human nouns, and non-human masculine plural nouns have the lowest accuracy rates. To more clearly understand the mechanism behind L2ers difficulty with these forms, below we look at the type of errors they make in production with respect to non-human plural nouns and dual nouns. This should allow us to better understand whether learners are indeed struggling with these forms due to featural mismatch, and if so, whether the mismatch in gender or number is more problematic.

4.3.1.4 Error Analysis

This section analyzes participants’ errors with respect to clitic production (gender and number). The first section reports errors involving incorrect gender and number specification on the clitic when the target noun is non-human plural. For both groups, participants make more gender errors (LP=76%, HP=52%) than number errors (LP=60%, HP=44.32%). Both groups provide masculine clitic responses instead of feminine ones.
(higher errors for LP group). Table 4.4 and Table 4.5 below provide the number of clitic responses by number and gender.

Table 4.4. Response Clitic Gender and Number for Non-Human Masculine Plural Target Nouns for LP

<table>
<thead>
<tr>
<th>Number/Gender</th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>22</td>
<td>10</td>
</tr>
<tr>
<td>Dual</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Plural</td>
<td>39</td>
<td>7</td>
</tr>
</tbody>
</table>

It is clear that for the LP group, when the target noun is nonhuman masculine plural, their preferred choice is a masculine clitic, but for the HP group there is no clear pattern.

Table 4.5. Response Clitic Gender and Number for Non-Human Masculine Plural Target Nouns for HP

<table>
<thead>
<tr>
<th>Number/Gender</th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>19</td>
<td>30</td>
</tr>
<tr>
<td>Dual</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Plural</td>
<td>27</td>
<td>12</td>
</tr>
</tbody>
</table>

When the target noun is non-human feminine plural, the percentage of gender and number errors decreases across both groups. For the LP group, participants switch the clitic to plural 38.75% of the time and dual 6.25% of the time. They only switched to masculine clitics 37.5% of their total clitic responses. For the HP group, their errors reduced almost by half. They used masculine clitics 26.13% and switched to plural only 28.41% of their total clitic responses.
Overall, in both proficiency groups, participants do not produce any dual clitics (except for the 6.25% reported above). Also, more gender and number errors are made when the target noun is non-human masculine plural. Since there is a 2-way mismatch in these cases, it is more problematic for these L2ers. Clitic agreement for non-human feminine plurals seems less problematic since the agreement mismatch only occurs on number.

We now turn to errors involving incorrect number specification on the clitic when the target noun is dual. Recall that with dual nouns the clitic must agree with the noun in number only. Table 4.6 and Table 4.7 show clitic number response for both proficiency groups when the target noun is dual masculine and dual feminine respectively.

| Table 4.6. Clitic Number Response when Target Noun Is Dual for LP Group |
|-----------------------------------------------|-|-|
| Response Number/Target Noun Gender | Masculine | Feminine |
| Singular | 53 | 77 |
| Dual | 43 (correct) | 43 (correct) |
| Plural | 64 | 40 |

| Table 4.7. Clitic Number Response when Target Noun Is Dual for HP Group |
|-----------------------------------------------|-|-|
| Response Number/Target Noun Gender | Masculine | Feminine |
| Singular | 56 | 45 |
| Dual | 77 | 104 |
| Plural | 43 | 27 |

The tables above indicate that participants use both singular and plural clitics where the dual form is required. Interestingly, this is in contrast with the results of
Experiment 2, where it was clear that the participants strongly preferred singular adjectives to modify dual nouns rather than plural ones.

4.3.1.5 Discussion

The results from the production task indicate that across both proficiency groups, L2ers were better with gender than with number agreement with DO clitics. However, morphological variability still remains a persistent problem for L2 learners even at advanced stages of proficiency (Franceschina, 2001; Lardiere, 1998 and White, 2003). The following are examples of gender errors in clitics (33 – 37) made by both LP and HP groups. The examples include a dialogue between the experimenter (E) and the participant (P). To control for animacy effects (human vs. non-human) all examples below are human singular masculine and feminine targets nouns.

*Participant LP8, Low L2 Arabic*

33. E: alwalad yashab albint. alwalad yashaba……. (The boy pulls the girl. The boy pulls ….)
   P: *yashab -hu (pulls – CL-MASC.SG)

*Participant HP5, Advanced L2 Arabic*

34. E: alwalad yadhan albint. alwalad yadhana……. (The boy paints the girl. The boy paints ….)
   P: *yadhan -hu (paints – CL-MASC.SG)

*Participant LP1, Low L2 Arabic*

35. E: albint tohib alwalad. albint tohibu ……. (The girl loves the boy. The girl loves ….)
   P: *tohibu-ha (loves – CL-FEM.SG)
In examples 33 and 34, the participants are using masculine clitics in feminine contexts. However, examples, 35 through 37 reveal the occurrence of feminine clitics in masculine contexts. Similar to results reported in Chapter 3, the emergence of feminine clitics in masculine contexts fails to support McCarthy’s (2007) MUSH which claims that when L2ers make gender or number errors they will invariably select the underspecified form i.e. masculine, singular. The results of this experiment like the previous one thus suggest that the underspecified form might not always be masculine in every language. This issue will be revisited in section 4.4. To show more examples of the use of feminine clitics in masculine contexts, sentences 38 through 40 below are examples of non-human masculine singular target nouns with feminine clitic production.

Participant LP6, Low L2 Arabic

38.  E: alwalad yashud alkalb. alwalad yashudu …….. (The boy pulls the dog (masc.sing. The boy pulls …..)
    P: *yashud -ha (pulls – CL-FEM.SG)

90
Participant *LP12, Low L2 Arabic*

39. E: alwalad yadfaʕ al kitaab. alwalad yadfaʕu ……. (The boy pushes the book
(masc.sing. The boy pushes ….)
P: *yadfaʕ -ha (pushes – CL-FEM.SG)

Participant *HP3, Advanced L2 Arabic*

40. E: Albint tushahid allʕaaʔir. Albint tushahid ……. (The girl sees the bird
(masc.sing. The girl sees ….)
P: *tushahid -ha (sees – CL-FEM.SG)

Similar to human nouns, the use of feminine singular clitics also emerges with non-human singular masculine nouns. These errors confirm that the use of feminine is attested in masculine contexts.

Number agreement also remains a problem for both low proficiency and advanced speakers. Singular contexts were largely unproblematic: for both human and non-human targets, error rates were less than 8%. Rather, the biggest issue with number variability is in duals and non-human plurals.

When looking at accuracy rates for human vs. non-human, we also find a similar pattern to the results in Chapter 3. The lowest accuracy rates are non-human masculine plural targets. LP participants are choosing between singular masculine and plural masculine clitics with no clear indication on their preferred choice.
The question remains why the dual presents a particular problem for learners. One potential explanation for low accuracy rates in dual is simply that this feature is absent in participants’ native language, English. However, these participants were clearly able to acquire gender, despite its absence from the L1.

On the face of it, Table 4.5 and Table 4.6 show that there are no clear error patterns emerging. However, the errors of dual clitics are systematic. Animacy seems to play a role in DO clitic production. Table 4.8 and Table 4.9 break down dual production for clitics with dual human compared to dual non-human for both groups. For human dual nouns, plural clitics are the prevailing choice when making errors. For non-human dual nouns, singular clitics are the prevailing choice for errors. If the same error patterns are attested in comprehension then we can safely assume that these L2ers are collapsing duality into plurality for clitics. This issue will be revisited in section 4.3.2.4.

**Table 4.8. Average Production of Clitic Number with Dual Human Nouns by Proficiency**

<table>
<thead>
<tr>
<th>Human Dual Nouns</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>11%</td>
<td>6%</td>
</tr>
<tr>
<td>Dual</td>
<td>17% (correct)</td>
<td>30% (correct)</td>
</tr>
<tr>
<td>Plural</td>
<td>23% (default)</td>
<td>15% (default)</td>
</tr>
</tbody>
</table>

**Table 4.9. Average Production of Clitic Number with Dual Non-Human Nouns by Proficiency**

<table>
<thead>
<tr>
<th>Non Human Dual Nouns</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>30% (default)</td>
<td>24% (default)</td>
</tr>
<tr>
<td>Dual</td>
<td>10% (correct)</td>
<td>22% (correct)</td>
</tr>
<tr>
<td>Plural</td>
<td>10%</td>
<td>6%</td>
</tr>
</tbody>
</table>
4.3.2 Comprehension Results

Accuracy rates were analyzed statistically with a series of mixed effects logistic regression models. Statistical significance for each fixed effect was assessed using likelihood ratio tests, a method of nested model comparison that determines whether model fit is significantly improved by adding additional fixed effects while controlling for added complexity. All analyses were conducting using R (R Core Team, 2013), and the lme4 package (Bates 2010). For each analysis, group is a between-subjects fixed effect with two levels (low, intermediate) gender is a within-subjects fixed effect with two levels (masculine or feminine gender), and number is a within-subject fixed effect with three levels (singular, dual or plural number). Where indicated, analyses may also include a fixed effect of animacy with two levels (human or non-human). All factors were sum coded, meaning individual levels of each factor were compared to the grand mean. Participant, target noun and target clitic are included as random effects for all models.

4.3.2.1 Gender Analysis

The analysis of gender included a total number of 4,320 tokens across both proficiency groups. Table 4.7 below shows the mean accuracy by proficiency group and gender. Overall, accuracy rates were higher for the HP group than the LP group. A mixed-effects model with proficiency and clitic gender as fixed effects confirmed a significant main effect of proficiency ($\chi^2(4)=44.01$, $p<0.0001$), but revealed no main effect of gender ($\chi^2(5)=2.32$, $p=0.13$). In addition, the model revealed a significant interaction between proficiency and gender ($\chi^2(6)=4.41$, $p=0.04$).
4.3.2.2 Number Analysis

For the analysis of number, 4,320 were analyzed. Figure 4.8 below shows mean accuracy rates by proficiency group and number. Overall, the HP had higher accuracy rates than LP. This was confirmed by a mixed-effects model with proficiency and clitic number as fixed effects, which revealed significant main effects for proficiency ($\chi^2(4)=38.54$, $p<0.0001$) and number ($\chi^2(6)=7.68$, $p=0.02$), but no significant interaction ($\chi^2(8)=1.72$, $p=0.42$). Individual comparisons revealed that for both proficiency groups, dual is significantly worse than singular (LP: dual-sg $\beta = -1.72 \pm 0.33$, $p<0.001$; and for HP: $\beta = -1.82 \pm 0.57$, $p=0.003$), and dual is significantly worse than plural for the LP group only ($\beta = -1.16 \pm 0.34$, $p=0.002$).

Figure 4.7. Mean % accuracy of clitic inflection on the comprehension task by proficiency and gender
Figure 4.8. Mean % accuracy of clitic inflection on the comprehension task by proficiency and number

Just as in the production task, we expect a decrement in accuracy for non-human plural nouns, where clitic agreement for gender and number are neutralized. If learners treat this as a featural mismatch, we would expect the decrement to specifically target masculine plural non-human target nouns. The cases of masculine non-human plural target nouns could be more problematic than feminine plural non-human nouns. In the masculine cases, the clitic mismatches in both gender and number, but in feminine cases the mismatch only occurs on number. This issue is discussed in the following section.

4.3.2.3 Further analysis of human vs. non-human

Figure 4.9 and Figure 4.10 show the mean percent accuracy of clitic number and gender inflection for singular, dual and plural and human, non-human target nouns for
each proficiency group respectively. Recall that when the target noun is non-human masculine or feminine plural the correct clitic takes the feminine singular form. Thus for non-human nouns, there are no cases in which the target clitic agreement is plural. For both proficiency groups, there is a clear decrement in performance for non-human singular feminine clitics. This is precisely as expected if mismatching agreement features presents a problem for learners.

![Mean % correct clitic choice for human vs. non-human target nouns for LP](image)

**Figure 4.9.** Mean % correct clitic choice for human vs. non-human target nouns for LP

**Figure 4.9** above shows that LP learners have trouble overall with dual and plural in both human and non-human categories. What is interesting to note is that accuracy rates on feminine clitic agreement are lower than for masculine clitics. Closer analysis
reveals that these low accuracy rates are driven by masculine plural non-human target nouns (38%) and feminine non-human plural target nouns (50%). By contrast, when the target noun itself is non-human feminine singular—i.e., when there is no feature mismatch—accuracy rates are much higher (67%).

Likelihood ratio tests for the LP group (Figure 4.9) were used to assess the effect of target clitic number, target clitic gender and animacy on accuracy. This confirmed a significant main effect of number ($\beta =0.51\pm0.83$, $p<0.0001$), a significant interaction of animacy and target clitic number ($\beta =0.38\pm0.11$, $p=0.0005$), but no significant effect for animacy ($\beta =0.28\pm0.76$, $p=0.7$) and no significant effect for target clitic gender ($\beta =0.11\pm0.79$, $p=0.15$). In summary, LP learners have lower accuracy rates on non-human noun targets overall, with a particularly low accuracy rate for masculine plurals in the non-human category.
Similar to the LP group, there is a general decrement in performance for non-human nouns in the HP group (Figure 4.10). The low accuracy rates on the singular non-human feminine target adjectival agreement are again driven largely by masculine non-human plural target nouns (52%). There is less of a problem with feminine non-human plural target nouns (78%). Likelihood ratio tests assessing the effect of target clitic number, target clitic gender and animacy on accuracy in this group confirmed a significant main effect of animacy ($\beta = 0.29 \pm 0.84$, $p=0.001$), a significant main effect of target clitic number ($\beta = 0.89 \pm 0.11$, $p<0.0001$), a significant interaction of target clitic number and animacy ($\beta = 0.62 \pm 0.18$, $p=0.0008$) and a marginal interaction of animacy, target clitic gender and target clitic number ($\beta = -0.31 \pm 0.18$, $p=0.09$). These results confirm lower accuracy rates on non-human noun targets overall, and with lower
accuracy rates on plural targets overall, though more so with masculine than feminine targets.

It is clear that for both proficiency groups dual non-human plural nouns are the most problematic. The low accuracy rates on non-human plurals are also observed in the production task. These results confirm that these L2ers are in fact treating the cases of non-human plural as featural mismatches between the noun and clitic; therefore having trouble identifying the correct picture to match the stimuli. The following section will investigate further the types of errors these L2ers make in comprehension.

4.3.2.4 Error Analysis

This section analyzes participants’ errors with respect to clitic comprehension (gender and number). The first section reports errors involving incorrect gender and number specification on the clitic when the target noun is non-human masculine plural. For both groups, participants make more gender errors (LP=62.78%, HP=48.98%) than number errors (LP=22.78%, HP=22.23%). The LP group provides more masculine clitics responses instead of feminine. Table 4.10 and Table 4.11 below provide the number of clitic choices by number and gender.

| Table 4.10. Clitic Number Choice when Target Noun Is Non-Human Masc Pl. for LP Group |
|-------------------------------|---------|---------|
| **Number/Gender**  | **Masculine** | **Feminine** |
| **Singular**         | 72 (default)  | 67 (correct)   |
| **Dual**             | 41        | 0         |
| **Plural**           | 0         | 0         |
Similar to the production task, when the target noun is non-human feminine plural, the percentage of gender and number errors decreases across both groups. Gender errors for the LP group are only 18.34% and 8.89% for the HP group. Number errors are slightly higher for the LP group, where almost half their choices (44.5%) are dual clitics. The HP group switches to dual only in 19.5% of their total responses.

We now turn to errors involving incorrect number specification on the clitic when the target noun is dual masculine and feminine for both groups. With dual nouns, the clitic must agree with the noun in number only. Table 4.12 and
Table 4.13 show clitic number choice for both proficiency groups when the target noun is dual masculine and dual feminine respectively.

Table 4.12. Clitic Number Response when Target Noun Is Dual for LP Group

<table>
<thead>
<tr>
<th>Response Number/Target Noun Gender</th>
<th>Masculine</th>
<th>Feminine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>155</td>
<td>142</td>
</tr>
<tr>
<td>Dual</td>
<td>152</td>
<td>165</td>
</tr>
<tr>
<td>Plural</td>
<td>53</td>
<td>53</td>
</tr>
</tbody>
</table>
From the tables above, participants in both proficiency groups erroneously use singular clitics where dual clitics are required. They rarely use plural clitics in place of duals. These results are interesting since (1) they look different from what we saw in the production task, and (2) we might have expected that participants would use plural clitics when making errors in the dual, since duality is collapsed with plural in their native language, i.e. English.

4.3.2.5 Discussion

Results in the comprehension task indicate that these L2ers were better able to identify gender agreement patterns than number agreement patterns with respect to DO clitics in Arabic. Results also indicate that morphological variability is a persistent problem even at advanced levels of proficiency (Franceschina, 2001; Lardiere, 1998 and White, 2003). When analyzing errors in the feature gender, feminine gender emerges in masculine contexts. L2ers selected feminine clitics when the context was masculine. These errors were also controlled for animacy (human vs. non-human) and similar patterns emerged for both environments.

By now we have established that the issue with number agreement stems from the subclass of nouns that are non-human plural specifically the masculine. In these cases, learners in both proficiency groups are choosing masculine singular and masculine dual
elitics instead of feminine singular elitics. This is also a persistent problem across all tasks.

Similar to the production task, dual remains a persistent problem for these L2ers with higher accuracy rates in the comprehension task. A closer look at the results of dual DO object clitic comprehension, Table 4.12 and
Table 4.13 show the incorrect number specification on the clitic when the target noun is dual. On the face of it, one might think that L2ers prefer singular number as their default choice. Similar to the production task, animacy effects were also found in DO clitic comprehension. Table 4.14 below shows that L2ers select the picture with the singular non-human noun with a dual clitic, but they never select the picture with the plural non-human noun. These results suggests that (1) the interlanguage of these L2ers is such that the feature dual is conflated with plural and (2) these L2ers are aware of Arabic’s agreement system between human and non-human where the former triggers full agreement features and the latter triggers singular feminine agreement on the clitics.

<table>
<thead>
<tr>
<th>Human Dual Nouns</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>14%</td>
<td>7%</td>
</tr>
<tr>
<td>Dual</td>
<td>22% (correct)</td>
<td>37% (correct)</td>
</tr>
<tr>
<td>Plural</td>
<td>15%</td>
<td>7%</td>
</tr>
</tbody>
</table>

Table 4.14. Average Clitic Choice on Comprehension Task with Human Dual Nouns by Proficiency

<table>
<thead>
<tr>
<th>Non Human Dual Nouns</th>
<th>LP</th>
<th>HP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Singular</td>
<td>28% (default)</td>
<td>17% (default)</td>
</tr>
<tr>
<td>Dual</td>
<td>23% (correct)</td>
<td>34% (correct)</td>
</tr>
<tr>
<td>Plural</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 4.15. Average Clitic Choice on Comprehension Task with Non-Human Dual Nouns by Proficiency

---

9 There does not seem to be a clear pattern of clitic number choice when the target noun is dual human
4.4 General Discussion

To answer the first research question: What are the general features of L2 speakers’ knowledge of gender and number agreement in Arabic clitic agreement? If we compare results from both production and comprehension we find that morphological variability is a persistent problem even at higher levels of proficiency. Gender features are present in these L2ers IL. Accuracy rates are slightly higher for gender in the comprehension task than the production task. On the other hand, accuracy rates on number features are significantly lower than gender for these L2 speakers. While singular number is unproblematic across all proficiency levels, dual was clearly problematic for learners in both tasks – higher in comprehension. Non-human plural targets remain problematic for both proficiency groups with non-human plural masculine being the most problematic category.

Turning to the second research questions in this chapter: What are the error patterns of agreement morphology in terms of gender (masculine, feminine) and number (singular, dual, plural) within DO clitic agreement? Starting with error patterns in gender agreement, the emergence of feminine in masculine contexts was attested in both experiments. These results lead us back to the same findings in Chapter 3 where L2ers also used feminine adjectives in masculine contexts. These results are problematic for McCarthy’s (2007) MUSH which claims that when L2ers make gender or number errors they will select the underspecified form i.e. masculine, singular. This suggests the possibility that the underspecified form in Arabic is not straightforwardly the masculine, but at least in some contexts is the feminine. In support of this claim, numerous L1 Arabic
studies have shown that children overgeneralize feminine gender and feminine plural forms (Moawad, 2006 and Omar, 1973). Other studies on heritage speakers of Arabic have shown similar findings, where these speakers have not fully acquired the complete Arabic number system and tend to use feminine plural forms as their default form (Albrini, Benmamoun & Chakrani, 2013 and Albrini & Benmamoun, 2014). In one Arabic L2 study by Alkhohlani (2016), she found that L2ers tend to overgeneralize feminine gender assignment on both masculine and feminine nouns. Finally, the agreement paradigm itself shows that in cases of gender neutralization, the feminine is used rather than the masculine.

Error patterns in dual DO clitics are similar across both comprehension and production. First, I have established that these L2ers are sensitive to semantic gender when producing or choosing a clitic in sections 4.3.1.5 and 4.3.2.5. L2ers are collapsing dual with the plural category. We can argue that this is due to lack of the feature in their L1. However, this behavior is also observed in native speakers of Arabic. In Chapter 2, Experiment 1 confirmed that native speakers of Arabic clearly prefer using plural adjectives with dual animate nouns but singular adjectives with dual inanimate nouns similar to the human/ non-human N-ADJ agreement paradigm. This suggests the possibility that dual is being lost in Arabic and assimilating to plurals (Ferguson, 1959).

Error patterns in non-human masculine plurals across both tasks in chapter 3 and chapter 4 show that L2ers tend to either produce or select the masculine choice. In principle, this could support McCarthy’s (2007) MUSH with the errors reflecting an underspecified masculine form. However, in light of the fact that masculine is not clearly
playing the role of the default here, we have argued that problems in this case can be attributed to L2ers simply producing or selecting the masculine to match the target noun features.

The third research question in this chapter: *Is morphological variability in the production task similar to morphological variability in the comprehension task?* The answer to this question will help us determine if morphological variability is a competence or representational issue. Gender and number errors decrease in comprehension for both proficiency groups. Even though accuracy rates are slightly higher for gender and number in the comprehension task, the error patterns are qualitatively similar. In gender errors, we have seen the emergence of feminine in masculine contexts in both comprehension and production. For number errors, dual presents a particular difficulty, and results show the effect of animacy on dual clitic production and comprehension. For example, human plural targets, plural DO objects were produced or chosen. Finally, cases of featural mismatch cause problems in both production and comprehension. These results support McCarthy’s (2008) claim that morphological variability is more than just a performance issue, but a problem with representing the features of gender and number.

Answering the last research question: *What is the role of animacy (human vs. non-human) in the acquisition of Arabic L2 gender and number features in DO clitics?* Similar to the findings in Experiment 2 in Chapter 3, results in production and comprehension confirm that L2 learners are sensitive to animacy when acquiring gender and number features on DO clitics. We can see that in both tasks, for both proficiency
groups, accuracy rates for human targets are numerically higher than non-human targets; with non-human masculine plural being the most problematic. Similar to results in Chapter 3, accuracy rates are higher in feminine non-human than masculine non-human targets across both groups. These observations lead to the same conclusions found in the previous chapter that L2 speakers are influenced by linguistic cues such as animacy when acquiring grammatical gender agreement features (Alarcon, 2010).

4.5 Conclusion

The results in the two tasks described in this chapter confirm the findings from chapters 2 and 3 that morphological variability will remain a persistent problem for L2 learners even at advanced levels of proficiency. Gender agreement in DO clitics is not problematic like number agreement especially with dual and non-human plurals. Dual number is more problematic in the DO clitic domain than the N+ADJ domain.

In addition, the use of feminine gender in masculine contexts and the overgeneralization of feminine in L2 Arabic suggest that the underspecified form in Arabic may be feminine. Finally, the effect of animacy (human vs. non-human) was present in both tasks; producing and interpreting correct gender and number was overall more difficult for non-human targets. Neutralized agreement likewise presented a persistent and substantial problem for learners across both tasks. Additional research is needed with non-human plural nouns not just with adjective agreement, and DO clitic agreement but also in other domains such as pronouns, demonstratives, and verbs. This suggestion will be taken up in the following chapter where it will discuss questions for further research.
Chapter 5. Conclusion

This dissertation presented original research on morphological variability in nominal morphology in L2 Arabic. I here summarize the results of the experiments described in this dissertation. Then I propose questions for future research.

5.1 Summary of Major Findings

This dissertation aims to explain L2 morphological variability and the variants involved in default morphology for an understudied language: Arabic. By looking at several different components of inflectional morphology across comprehension and production, it is clear that substitution errors are principled and not random. In addition, data in this dissertation shows that underspecified features do not necessarily correspond to the unmarked values i.e masculine, singular. The data shows that the underspecified form for gender for L2 Arabic learners might be feminine.

The first set of experiments (reported in chapter 2) was a pilot study of morphological variability in gender and number agreement in L2 Arabic NPs. The first experiment was a study of elicited production N + ADJ combinations in singular and dual number only. The second experiment tested comprehension of N + ADJ combinations in singular and dual numbers as well. Results in production revealed overall decrease in variability with increasing proficiency. Results also showed that dual seems problematic for both proficiency groups and for native speakers. Comprehension accuracy rates were
near ceiling for all groups. The asymmetries in results lead me initially to support predictions by the MSIH. However, two issues needed to be addressed before supporting the MSIH. The first issue was that native speakers clearly did not treat dual as “two” but classified the dual as plural – anything more than one. They assigned gender and number on the adjective based on the animacy of the dual noun. For animate dual nouns, they used the human plural agreement paradigm, and for inanimate dual nouns, they used the non-human plural agreement paradigm. The second issue was the methodology. The stimuli between comprehension and production were different. In particular, the production task used non-human inanimate objects, and in the comprehension task, the pictures were all human. Therefore, the asymmetry in stimuli between production and comprehension attributed to this difference.

The second set of experiments (reported in chapter 3) built on the pilot study in chapter 2. The first experiment was a study of elicited production N + ADJ combinations in singular, dual and plural number with animacy being manipulated. The second experiment tested comprehension of N + ADJ combinations in singular, dual and plural number with animacy being manipulated as well. In both experiments, feminine gender was attested in masculine contexts, suggesting it is the underspecified form. Human accuracy rates were overall numerically higher than non-human accuracy rates across both groups. Morphological variability is persistent even at advanced levels of proficiency. Finally, morphological variability extends to comprehension and is qualitatively similar to production, supporting McCarthy (2008) that this variability is a representational issue.
The third set of experiments (reported in chapter 4) presented original data on nominal morphology in L2 Arabic specifically the features if gender and number in DO clitics. The first experiment was elicited production of gender and number agreement in DO clitics in singular, dual and plural number with animacy being manipulated. The second experiment was comprehension of gender and number agreement in DO clitics in singular, dual and plural number with animacy being manipulated. Again, in both experiments, feminine gender was attested in masculine contexts, suggesting it is the underspecified form. Human accuracy rates were overall numerically higher than non-human accuracy rates across both groups. Dual errors in both experiments were affected by animacy of the target noun question. L2ers similar to natives were sensitive to these cues when producing or selecting the DO clitic, suggesting that L2ers were collapsing dual with plural. Finally, morphological variability extends to comprehension and is qualitatively similar to production, supporting McCarthy (2008) that this variability is a representational issue.

5.2 Questions for Future Research

Morphological variability has been widely investigated in many languages but only a few studies have been devoted to investigating morphological variability in L2 Arabic. Bearing in mind the richness of Arabic morphology and syntax in general, and gender and number in particular, future research on the acquisition of gender and number not just in the nominal domain but in the verbal domain might bring more insight into the field of second language acquisition.
First, the data collected in this dissertation was mainly done through elicited production. An elicited production method is a robust method to target the specific structure tested. However, collecting spontaneous production samples would allow the researcher to investigate more structures in the L2. The researcher would not be limited to N+ADJ combinations and DO clitics. Structures like subject + verb agreement, pronouns, number + noun agreement, grammatical case (nominative vs. accusative), sound plurals versus broken plurals, specifically testing the effects of overt morphology on gender agreement, and many more can be tested and analyzed.

Another direction for future research would be to confirm further my findings that feminine is the underspecified form in Arabic. My data has shown that feminine is attested in several masculine contexts, but these were all in the nominal domain. Future research in verbal domain would support my claim especially that Arabic has full agreement features between subject and verbs in SVO word order and gender agreement in VSO word order.

Finally, another interesting direction for future research would be to further consider dual number in L2 acquisition. Even though L2 learners are explicitly learning these structures in the classroom, their agreement assignment of features tells another story. These L2ers are clearly treating dual as plural similar to native speakers as we have seen in chapter 4. A study solely focused on comparing dual production and comprehension with animacy effects in native speakers and L2ers would help us better understand the mechanism of dual acquisition in L2 Arabic.
Appendix A: Demographic Questions

1. What is your native language?
2. Do you speak any other languages natively?
3. As an adult, have you learned any other languages?
4. When did you start learning Arabic? (age of onset)
5. How many Arabic courses have you studied so far at the university level?
6. Have you resided in an Arabic speaking country? If yes, how long?
7. Are you studying Arabic now? If yes, which level are you placed in?
# Appendix B: Vocabulary Set 1 and 2 for Production and Comprehension in NPs

## Experiment

### Vocabulary Set 1

<table>
<thead>
<tr>
<th>Human</th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masc</td>
<td>ustaad</td>
<td>raqisa</td>
<td>asate (\text{\textit{d}})a</td>
</tr>
<tr>
<td>Fem</td>
<td>2</td>
<td>2</td>
<td>2 (\text{\textit{d}})ancers</td>
</tr>
<tr>
<td>Mas (\text{\textit{c}})c</td>
<td>ustaad</td>
<td>raqisa</td>
<td>asate (\text{\textit{d}})a</td>
</tr>
<tr>
<td>Mas (\text{\textit{c}})c</td>
<td>raqisa</td>
<td>2</td>
<td>2 (\text{\textit{d}})ancers</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-human</th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masc</td>
<td>qalam</td>
<td>sayyara</td>
<td>aqlam</td>
</tr>
<tr>
<td>Fem</td>
<td>sayyara</td>
<td>2</td>
<td>2 (\text{\textit{c}})ars</td>
</tr>
<tr>
<td>Masc</td>
<td>kitaab</td>
<td>2</td>
<td>2 (\text{\textit{b}})uildings</td>
</tr>
<tr>
<td>Fem</td>
<td>kitaab</td>
<td>2</td>
<td>2 (\text{\textit{b}})uildings</td>
</tr>
<tr>
<td>Masc</td>
<td>kalb</td>
<td>2</td>
<td>2 (\text{\textit{c}})ats</td>
</tr>
<tr>
<td>Fem</td>
<td>kalb</td>
<td>2</td>
<td>2 (\text{\textit{c}})ats</td>
</tr>
<tr>
<td>Masc</td>
<td>hisaan</td>
<td>baqara</td>
<td>ahsena</td>
</tr>
<tr>
<td>Fem</td>
<td>baqara</td>
<td>2</td>
<td>2  (\text{\textit{c}})ows</td>
</tr>
</tbody>
</table>

### Examples

- Human:
  - 'doctor': t'\(\text{\textit{b}}\)eeb
  - 'employee': mowa\(\text{\textit{d}}\)af\(\text{\textit{a}}\)
- Non-human:
  - 'pen': qalam
  - 'building': kitaab
  - 'horse': hisaan
<table>
<thead>
<tr>
<th>Human</th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>Masc</td>
<td>δʾaabit</td>
<td>ḍʾaabitayn</td>
<td>‘police officer’</td>
</tr>
<tr>
<td>Masc</td>
<td>momaridʾa</td>
<td>momaridʾataan</td>
<td>‘nurse’</td>
</tr>
<tr>
<td>Masc</td>
<td>dahin</td>
<td>dahinayn</td>
<td>‘2 police officers’</td>
</tr>
<tr>
<td>Masc</td>
<td>mowṣʾawira</td>
<td>mowṣʾawiratayn</td>
<td>‘photographer’</td>
</tr>
<tr>
<td>Masc</td>
<td>tʿabbāχ</td>
<td>tʿabbāya</td>
<td>‘2 chefs’</td>
</tr>
<tr>
<td>Masc</td>
<td>laʾśiba</td>
<td>laʾśibatayn</td>
<td>‘player’</td>
</tr>
<tr>
<td>Masc</td>
<td>mokanekeya</td>
<td>mokanekeyatayn</td>
<td>‘mechanic’</td>
</tr>
<tr>
<td>Masc</td>
<td>tʿabbāχuun</td>
<td>tʿabbāχuun</td>
<td>‘2 painters’</td>
</tr>
<tr>
<td>Masc</td>
<td>moẓare</td>
<td>moẓareṣayn</td>
<td>‘farmer’</td>
</tr>
<tr>
<td>Masc</td>
<td>moẓareṣuun</td>
<td>moẓareṣuun</td>
<td>‘mechanic’</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Non-human</th>
<th>Singular</th>
<th>Dual</th>
<th>Plural</th>
</tr>
</thead>
<tbody>
<tr>
<td>qalam</td>
<td>qalamayn</td>
<td>aqlam</td>
<td>‘desk’</td>
</tr>
<tr>
<td>qalam</td>
<td>sayyaratayn</td>
<td>sayyarat</td>
<td>‘desk’</td>
</tr>
<tr>
<td>kitaab</td>
<td>kitaabayn</td>
<td>kutub</td>
<td>‘phone’</td>
</tr>
<tr>
<td>kitaab</td>
<td>binayat</td>
<td>binaya</td>
<td>‘bookshelf’</td>
</tr>
<tr>
<td>kalb</td>
<td>kalbayn</td>
<td>kilaab</td>
<td>‘rabbit’</td>
</tr>
<tr>
<td>kalb</td>
<td>qittatayn</td>
<td>qittat</td>
<td>‘camel’</td>
</tr>
<tr>
<td>hisaan</td>
<td>hisaanayn</td>
<td>ahsena</td>
<td>‘lion’</td>
</tr>
<tr>
<td>hisaan</td>
<td>baqaratayn</td>
<td>baqaraat</td>
<td>‘turtle’</td>
</tr>
</tbody>
</table>
Appendix C: Vocabulary Set 1 and 2 for Production and Comprehension in DO

Clitics Experiment

Human Trials:

Singular

1. He pushed the student (masc) – hu
2. He sees the teacher (m) – hu
3. He pulls the girl (fem) – ha
4. He paints the student (fem) – ha
5. She loves the boy (m) – hu
6. She see the doctor (m) – hu
7. She pushes the doctor (fem) – ha
8. She loves the teacher (fem) – ha

Dual

1. He pushed the students (masc) – huma
2. He sees the teachers (m) – huma
3. He pulls the girls (fem) – huma
4. He paints the students (fem) – huma
5. She loves the boys (m) – huma
6. She see the doctors (m) – huma
7. She pushes the doctors (fem) – huma
8. She loves the teachers (fem) – huma

Plural

1. He pushed the students (masc) – hum
2. He sees the teachers (m) – hum
3. He pulls the girls (fem) – hunn
4. He paints the students (fem) – hunn
5. She loves the boys (m) – hum
6. She see the doctors (m) – hum
7. She pushes the doctors (fem) – hunn
8. She loves the teachers (fem) – hunn
Non-Human Trials:

Singular

1. He pushed the book (masc) – hu
2. He pushed the table (fem) – ha
3. He pulls the dog (masc) – hu
4. He watches – sees the duck (fem) – ha
5. She loves the house (masc) – hu
6. She loves the car (fem) – ha
7. She watches the bird (masc) – hu
8. She loves the cat (fem) – ha

Dual

1. He pushed the books (masc) – huma
2. He pushed the tables (fem) – huma
3. He pulls the dogs (masc) – huma
4. He watches – sees the ducks (fem) – huma
5. She loves the houses (masc) – huma
6. She loves the cars (fem) – huma
7. She watches the birds (masc) – huma
8. She loves the cats (fem) – huma

Plural

1. He pushed the books (masc) – ha
2. He pushed the tables (fem) – ha
3. He pulls the dogs (masc) – ha
4. He watches – sees the ducks (fem) – ha
5. She loves the houses (masc) – ha
6. She loves the cars (fem) – ha
7. She watches the birds (masc) – ha
8. She loves the cats (fem) – ha
References


Biography

Baraa Rajab holds a Bachelor of Arts in English from the King Abdulaziz University, a Masters of Arts in Library and Information Science from King Abdulaziz University and a Masters of Arts in English with a concentration in Linguistics from George Mason University.