Age Effects on the Acquisition of Uninterpretable Features by Proficient Saudi Arabic Speakers of English*

Kholoud Al-Thubaiti
University of Essex

A recent account of fossilization in adult second language (L2) grammars is the ‘Interpretability Hypothesis’ (Hawkins & Hattori, 2006; Tsimpli & Dimitrakopoulou, 2007). It proposes that properties associated with uninterpretable features not already activated in the first language (L1) grammar will pose a learnability problem for older L2 speakers because they are inaccessible beyond a critical period (CP). The present study tests this hypothesis by examining the effect of age on the knowledge shown by proficient Saudi Arabic speakers of L2 English of two subtle linguistic properties associated with uninterpretable features: (i) the Gap Strategy in wh-interrogatives, and (ii) Reflexive Binding. While the former is differently instantiated in Arabic and English, the latter is similarly present in both languages. Using an acceptability preference task, results show that fossilization is selective, and is a reflex of L1–L2 grammatical differences. As predicted, the advanced adult starters showed a persistent L1 effect (de-learning problem) in the acquisition of the gap strategy in wh-interrogatives, but hardly had any problems with reflexive binding. Thus, it is argued that convincing evidence to refute a maturation account should be drawn from testing contrasting structures rather than similar ones. As for the child starters, while the high-advanced group showed native-like competence in all the tested constructions, the advanced group did not. Divergence in the child starters' L2 grammar might be the effect of other factors like proficiency.

1 INTRODUCTION

Selective fossilization in adult second language (L2) grammars of varied first language (L1) backgrounds is an appealing topic in L2 acquisition research. The ‘Interpretability Hypothesis’ (Hawkins & Hattori, 2006; Tsimpli & Dimitrakopoulou, 2007) is a recent account for this phenomenon. In this account, L1 and age effects are intertwined; fossilization is predictable on the basis of whether the target property is related to uninterpretable/interpretable features, and their status in L1 compared to L2. The locus of the learnability problem is in the narrow syntax; uninterpretable features are subject to maturation, and thus are candidates for persistent divergence in the grammars of L2 late starters if not already activated in their L1. Interpretable features, on the other hand, remain accessible throughout life, rendering them unproblematic for late starters even when not activated in their L1.

Tsimpli and Dimitrakopoulou (2007), Hawkins et al. (2007), and Hawkins and Hattori (2006) are very recent empirical studies that report on divergent representations where an L2 property is associated with uninterpretable features not activated in L1 grammar. In these studies, (highly) proficient post-critical period L2 learners of English were targeted. Tsimpli and Dimitrakopoulou examined intermediate and advanced L1 Greek speakers on their knowledge of acceptable gaps vs. unacceptable resumptives in English wh-interrogatives. Testing aspectual interpretation, Hawkins et al. looked at advanced L1 speakers of Chinese, Japanese, and a mixture of verb-raising languages. Hawkins and Hattori examined how multiple wh-questions are interpreted by highly proficient L1 Japanese. These studies together

* I am greatly indebted to Roger Hawkins for his useful comments and suggestions. Needless to say any shortcomings are my responsibility.
suggest that a maturational account that integrates L1 transfer offers a principled understanding of selective fossilization in adult L2 acquisition.

Following up current work on the ‘Interpretability Hypothesis’, this study is carried out with (highly) proficient L1 Saudi Arabic-speakers of English. The aim of this paper is twofold: (i) to evaluate age effects by comparing the performance of two age groups (early and late starters), and (ii) to evaluate L1 effects by testing two properties where one is differently instantiated (resumptives vs. gaps), and the other is similarly established (reflexive binding) in L1 and L2. This is to argue that strong evidence to refute a maturational account should be based on contrasting properties rather than similar ones because success in the latter is confounded with positive transfer from the L1. The paper is organized as follows: section 2 reports in detail the empirical study and its findings. Section 3 discusses the results in light of the research questions, and section 4 concludes the argument.

2 The Current Study

Addressing the issue of selective fossilization, two subtle linguistic properties are considered where one is differently instantiated (i.e., gap vs. resumptives) in Saudi Arabic (henceforth, SA) and English, and the other is similarly established (i.e., reflexive binding). The rationale for this design is to argue that age effects are likely to appear with contrasting properties rather than similar ones. This is apparently due to age effects being confounded with L1 positive transfer, and thus, age effects are concealed. With contrasting structures, on the other hand, age effects are revealed through (persistent) divergence from natives' grammar due to L1 negative transfer. Evidence in both conditions suggests reliance on L1 grammar, but in the first condition it can be misleading when used to argue against maturational claims. In a recent critique of empirical research on age effects, Long (2005) strongly recommends testing typologically distant L1-L2 pairings. A maturational account is hence falsifiable if evidence drawn from parameterised structures show convergence on the natives' grammar. Comparing L2 speakers' knowledge of contrasting vs. similar properties could help in understanding the nature of the problem of fossilization.

Formation of wh-interrogatives is one parameterised aspect that distinguishes SA from English. While English employs the gap strategy exclusively (compare (1) with (2) where an asterisk indicates ungrammaticality throughout all examples), SA employs the resumptive strategy along with the gap strategy as two alternating options as in (3) and (4), respectively. According to Tsimpli and Dimitrakopoulou (2007), resumptive pronouns are overt realizations of [tlnfl] features at Logical Form (LF).

(1)  \textbf{Who}_i \text{ do you think } \textbf{Ø}_i \text{ will marry Susan?}
(2)  \textbf{Who} \text{ do you think (he) will marry Susan?}
(3)  Miin_i \text{ ti-hassb-i-h}_i \text{ räh yì-twawj Susan?}

Who 2FS-think-you-he will 3MS-marry Susan

‘Who do you think he will marry Susan?’
(4)  Miin_i \text{ ti-hassb-i } \textbf{Ø}_i \text{ räh yì-twawj Susan?}

Who 2FS-think-you Ø will 3MS-marry Susan

‘Who do you think will marry Susan?’

Reflexive binding, on the other hand, is similarly instantiated in SA and English. In both languages, reflexives are morphologically complex, and thus, they must be locally bound as in
According to Hicks (2005), locally bound reflexives encode \([u\text{Ref}]\) that matches with an antecedent bearing an \([i\text{Ref}]\).

\[\text{(5) TP[ Bill said that TP[John loves himself*-u]]].}\]

\[\text{(6) TP[ Bill qal innu TP[John yi-hub nafs-u-i]].}\]

\[\text{Bill said that John 3MS-love self-him}\]

‘Bill said that John loves himself.’

2.1 Research Questions

1. Will Saudi Arabic-speakers of different starting ages continue using resumptive pronouns along with gaps in English, or will they learn that only the gap strategy is applied?
2. Will Saudi Arabic-speakers of different starting ages realize that reflexives in English are similar to Arabic in that they must be locally bound?

If the ‘Interpretability Hypothesis’ holds true, resumptives will cause a de-learning problem for adult starters, and reflexive binding will be successfully acquired by both child and adult starters.

2.2 Tasks

A shortened version of Tsimpli and Dimitrakopoulou’s (2007) acceptability preference test was used to evaluate tolerance of resumptive pronouns. The test consisted of subject/object wh-interrogatives, some of which are ungrammatical by virtue of having resumptive pronouns, and some are grammatical with gaps. From a list of 30 items, 20 items were selected for testing, and 6 distracters were added with 4 grammatical and 2 ungrammatical to achieve a balanced set of grammatical and ungrammatical items. The test-items fall in five categories with 4 tokens for each (see 7 & 8). The reliability coefficient of alpha obtained is (.71) for grammatical wh-interrogatives, and (.87) for ungrammatical items.

\[\text{(7) Subject Extraction}\]

(a) Which athlete does John think can win the Olympics? (-that, -RP)
(b) Which party does John think \(*\text{it} was very popular?\) (-that, +RP)
(c) Who have you suggested \(*\text{that he} should not resign?\) (+that, +RP)

\[\text{(8) Object Extraction}\]

(a) Which parcel did you say that Mary sent yesterday? (+that, -RP)
(b) What did you say that Maria forgot \(*\text{it} when she was leaving home?\) (+that, +RP)

Testing knowledge of reflexives’ binding domain, White et al’s (1997) truth-value judgment task was adapted into the form of a story acceptability preference task. For each story, a reflexive is contrasted with a pronoun where possible readings are primed by context. Sentences were biclausal with finite and non-finite embedded clauses. For local antecedent readings, sentences with reflexives are appropriate, but ones with pronouns are inappropriate. The reverse pattern is given with non-local antecedent readings. There were three tokens for each condition. The test consisted of 12 stories and 4 story distracters. The reliability coefficient of alpha obtained is (.70) for appropriate continuations, and (.77) for inappropriate continuations. A sample story is given in (9) below.
Killer Harry was free again. Bill was very scared. Bill called a policeman so the policeman could guard him and make sure he was safe from Killer Harry.

(a) Bill asked the policeman to protect himself.
(b) Bill asked the policeman to protect him.

For both acceptability tasks, a 5-point rating scale (completely impossible -2 -1 0 +1 +2 completely possible) was adopted where the end-points of the scale express a robust acceptance or rejection, and (0) is taken to denote uncertainty.

2.3 Participants

A total of 25 (highly) proficient L2 Saudi Arabic-speakers of English formed the experimental group, in addition to 6 British English controls (henceforth, NS). The experimental participants were of high-advanced and advanced proficiency in English according to the Quick Oxford Placement Test (QOPT). Age at first significant immersion to English was applied as a criterion for qualifying to be part of the child starters group (n=14), and adult starters group (n=11). Based on L2 proficiency and age at first exposure (thereafter, AFE), the experimental group was divided into three subgroups as follows: (i) high-advanced child starters (henceforth, HA-Child) (n=8, mean AFE=3.5 yrs), (ii) advanced child starters (henceforth, A-Child) (n=6, mean AFE=3.17 yrs), and (iii) advanced adult starters (henceforth, A-Adult) (n=11, mean AFE=17.5 yrs). The experimental participants were university undergraduates and graduates majoring in English Linguistics and Literature in Saudi Arabia at the time of testing. It is important to point out that the child starters had lived in an English-speaking country, whereas all the adult starters except for one had not. However, in this study, immersion in a foreign language country is assumed to be relatively comparable to that in an English-speaking country, especially when the participants are mainly English majors.

2.4 Results

The average ratings of acceptance and rejection obtained by the participants were submitted to a series of (1 × 4) between-group analysis of variance (ANOVA). Results of both tasks on resumptives and reflexives are presented in the following sections.

2.4.1 Resumptives vs. Gaps in wh-interrogatives

In Table 1, results show a significant group effect on rejection ratings of subject resumptive pronouns in both conditions where the complementizer is present and absent; a stronger significant effect is shown with the former condition.1 With object resumptives, a trend for group effect approaches significance.

In all three conditions, high-advanced and advanced child starters show stronger rejection ratings for resumptive pronouns than advanced adult starters. Most importantly, the latter group inaccurately accepted subject resumptives in wh-interrogatives with a complementizer (see Table 1). Results of Tukey post-hoc comparisons show that the ratings of advanced adult starters significantly lag behind that of high-advanced child starters on the one hand, \(p<.05\), and natives, on the other, \(p<.01\), in the case of subject wh-interrogatives with and without a complementizer. Also, it is found that the ratings of advanced child starters

1 This construction involves violation of two parameterised aspects: that-trace effect due to the Empty Category Principle (ECP), and use of subject resumptives.
significantly lag behind that of natives in the case of subject wh-interrogative with a complementizer, \( p < .01 \).

<table>
<thead>
<tr>
<th>Tested Conditions</th>
<th>NS ((n=6))</th>
<th>HA-Child ((n=8))</th>
<th>A-Child ((n=6))</th>
<th>A- Adult ((n=11))</th>
<th>( df )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>*Subj. RP (-that)</td>
<td>-1.75 ((0.32))</td>
<td>-1.38 ((0.88))</td>
<td>-0.71 ((0.91))</td>
<td>-0.11 ((1.08))</td>
<td>3.27</td>
<td>5.443**</td>
</tr>
<tr>
<td>*Subj. RP (+that)</td>
<td>-1.67 ((0.54))</td>
<td>-1.13 ((0.65))</td>
<td>-0.04 ((0.60))</td>
<td>0.18 ((1.00))</td>
<td>3.27</td>
<td>9.716***</td>
</tr>
<tr>
<td>*Obj. RP (+that)</td>
<td>-1.46 ((0.64))</td>
<td>-1.41 ((0.77))</td>
<td>-1.00 ((0.63))</td>
<td>-0.59 ((0.74))</td>
<td>3.27</td>
<td>2.8568</td>
</tr>
</tbody>
</table>

\( p < .01, ** p < .001, p = .056 \). Arrows show sig. at \( p < .05 \) in the Tukey post-hoc tests.

Table 1

Rejection Mean Ratings of Ungrammatical wh-interrogatives (Resumptive Strategy)

Regarding the acceptance ratings on grammatical wh-interrogatives with gaps, ANOVA results yield no significant group effect with subject questions, but a significant effect with object questions. As shown in Table 2, with subject questions, the mean ratings obtained by all groups are fairly close. However, with object questions, advanced child and adult starters show the least acceptance ratings compared to high-advanced child starters and natives. Dunnett's post-hoc comparisons yield a marginally non-significant difference between advanced adult starters and high-advanced child starters groups, \( p = .056 \), even though the mean difference is visually quite big. A significant difference is shown between the acceptance ratings of advanced adult starters and natives, \( p < .05 \). The advanced child starters gave the weakest acceptance ratings on grammatical object wh-questions which nevertheless does not reach statistical significance with any of the groups².

<table>
<thead>
<tr>
<th>Tested Conditions</th>
<th>NS ((n=6))</th>
<th>HA-Child ((n=8))</th>
<th>A-Child ((n=6))</th>
<th>A- Adult ((n=11))</th>
<th>( df )</th>
<th>( F )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Subj. Gap (-that)</td>
<td>1.80 ((0.33))</td>
<td>1.63 ((0.38))</td>
<td>1.29 ((0.56))</td>
<td>1.23 ((0.69))</td>
<td>3.26</td>
<td>1.727</td>
</tr>
<tr>
<td>Obj. Gap (+that)</td>
<td>1.88 ((0.14))</td>
<td>1.75 ((0.27))</td>
<td>0.38 ((1.16))</td>
<td>0.88 ((0.85))</td>
<td>3.26</td>
<td>6.443**</td>
</tr>
</tbody>
</table>

\( p < .01 \). Arrows show sig. at \( p < .05 \) in the Dunnett post-hoc tests.

Table 2

Acceptance Mean Ratings of Grammatical wh-interrogatives (Gap Strategy)

2.4.2 Reflexive Binding

Results yield no significant effect due to group, \( p > .05 \), in both clause types (finite/non-finite); all groups strongly rejected inappropriate interpretations with reflexives binding to non-local antecedents, and accepted appropriate interpretations where reflexives bind to local antecedents. Appropriate interpretations with pronouns binding to non-local antecedents were also strongly accepted in finite and non-finite clauses by all groups. As shown in Table 3, the average mean ratings of all Saudi Arabic groups regardless of starting age fall within the range of the natives.

² This is possibly because the Standard Deviation \((1.16)\) indicates a lot of variation among the six participants in this group. In such a condition, caution is required in interpreting the obtained Mean.
3 DISCUSSION

The first question this paper tried to answer is whether Saudi Arabic-speakers of different starting ages will tolerate resumptives in English wh-interrogatives along with gaps. Recall that resumptives are overt realizations of uninterpretable features encoded in the Arabic L1 lexicon, but not part of the English L2 lexicon, and thus are predictable candidates for fossilization for adult L2 starters in a maturational account. To answer this question, two sets of data are considered: (i) rejection ratings of ungrammatical constructions with resumptives, and (ii) acceptance ratings of grammatical constructions with gaps. To start with rejection ratings, ungrammatical subject/object wh-interrogatives with resumptives are grammatically accepted in Arabic, and thus, they are taken as a measure of L1 effect. Low rejection ratings suggest divergence from the natives' grammar and reliance on L1, and high rejection ratings indicate convergence. As predicted, the advanced adult starters group showed the lowest rejection ratings for two ungrammatical constructions: subject wh-interrogatives (-that), and object wh-interrogatives (+that); they also inaccurately accepted ungrammatical subject wh-interrogatives (+that) (e.g., who have you suggested (*that he) should not resign?). The child starters groups, on the other hand, showed higher rejection ratings with the high advanced group outperforming the advanced group in all three tested constructions. As far as subject wh-interrogative (+that) constructions are concerned, the ratings of advanced L2 child and adult starters groups show divergence from the natives’ grammar. While the latter group showed acceptance, the former revealed uncertainty in their judgments. Tentatively, I argue that since both groups showed an asymmetrical rating performance on such ungrammatical constructions, different types of ratings (positive vs. negative) - even if very low and close to not sure - should have stemmed from different underlying representations. Arabic, unlike English, resorts to the resumptive strategy to avoid ECP violations (Mohammad, 2000). Since the advanced L2 adult starters mean rating is positive (acceptance), it is possible that they are...
analysing such constructions in light of their L1 grammar, suggesting a persistent L1 effect. In the face of that, the advanced child starters mean rating is negative (rejection), and thus, it is less likely to indicate reliance on L1 grammar. Such asymmetrical behaviour merits further investigation.

Turning to the acceptance ratings, subject/object *wh*-interrogatives with gaps are grammatical in Arabic and English. Therefore, these constructions should not be problematic for Saudi Arabic-speakers of any age group. However, a subject/object asymmetry is shown in the ratings of advanced adult starters, and advanced child starters only. This is based on their high acceptance ratings which are indistinguishable from that of natives on subject questions, and low acceptance ratings that lag behind natives on object questions. While subject questions are successfully acquired by Saudi Arabic-speakers of both age groups, object questions turn out to be problematic for the advanced adult group, and possibly the advanced child group. Caution, however, is warranted in interpreting the ratings of the latter group due to the high standard deviation obtained. Such asymmetrical performance is mysterious given that gaps alternate optionally with resumptives in Arabic grammar. This requires further investigation of the distribution of gaps and resumptives in subject and object questions. Assuming a persistent L1 effect, the low acceptance ratings of gaps in object questions possibly indicate that a resumptive pronoun is preferable over a gap in Arabic. However, this is a speculation and requires further testing.

The second question of this paper aimed at testing L1 positive transfer by evaluating how successful Saudi Arabic-speakers of different age groups will be in acquiring an L2 property like reflexive binding which is similarly instantiated in their L1 grammar. Recall that locally bound reflexives encode a \([u_{Ref}]\), and since they are already encoded in L1, it will be easily acquired by both age groups. As predicted, all (highly) proficient Saudi Arabic-speaker groups regardless of starting age show knowledge that reflexives must be locally bound, whereas pronouns can be non-locally bound. They strongly rejected non-local antecedents to bind with reflexives. Their ratings are indistinguishable from that of natives. All groups realise the differences between reflexives and pronouns in terms of their binding domain restrictions.

Based on this, I argue that these findings are consistent with predictions of the ‘Interpretability Hypothesis’, and thus a maturational account, for two main reasons: (i) adult starters did not converge on native grammars in any of the tested constructions that encode parameterised uninterpretable features (resumptives), and (ii) adult starters did not outperform child starters in any of these constructions. The fact that advanced child starters did not show convergence on all tested constructions does not conflict with a maturational account. Long (2005:290), in this regard, argues that within a maturational account it is not assumed that early starters will definitely converge on native-like competence; but the claim is that early starters can converge, and late starters cannot converge. Therefore, evidence for late starters attaining native-like competence could falsify a maturational account. However, this leaves us with some open questions for further research: why is convergence not inevitable in the case of all L2 child starters? Since it is generally accepted that Universal Grammar (UG) constrains child language acquisition, what factors can lead to divergence in child L2 grammar? Is it proficiency or L1 dominance? To answer these questions fairly, there should have been a high-advanced adult starters group in the study sample to compare their performance with the high-advanced child starters and advanced adult starters groups, but this group was not feasibly available.

4 CONCLUSION

To conclude, fossilization is selective, and is a reflex of L1-L2 grammatical differences. For L2 adult starters, a persistent L1 effect is inevitable when the L1 property (resumptives) is
associated with uninterpretable features encoded in the L1 lexicon, but not in the L2. On the other hand, where an L2 property (reflexive binding) is associated with uninterpretable features encoded similarly in L1 and L2, it can be successfully acquired by all age groups. Based on this, evidence arguing against a maturational account should not be extracted from testing properties similarly instantiated in L1 and L2; they should rather be drawn from contrasting structures. Divergence in the L2 child starters’ grammar is not typical as in the case of L2 adult starters; based on the study’s findings, advanced L2 child starters can possibly show divergent properties, but not the very advanced ones.

REFERENCES


Kholoud Al-Thubaiti

Language and Linguistics Department
Wivenhoe Park
University of Essex
Colchester
CO4 3SQ
United Kingdom
kalthu@essex.ac.uk