ON-LINE PROCESSING OF SENTENCES INVOLVING REFLEXIVE AND NON-REFLEXIVE PRONOUNS IN L1 AND L2 CHILDREN

THEODOROS MARINIS

1. Introduction

Research in adult second language (L2) acquisition has established systematic differences between L1 and L2 grammars even in L2 learners with near-native proficiency (Hawkins, 2001). In addition, recent studies of sentence processing by L2 adults have revealed systematic quantitative and qualitative differences between L2 adults, L1 adults, and L1 children (Juffs & Harrington, 1995; Marinis, Roberts, Felser, & Clahsen, 2005).

Compared to the substantial literature on adult L2 acquisition, relatively fewer studies have investigated child L2 acquisition. These have focused primarily on language production and have revealed among others that L2 children have difficulties with grammatical morphemes (Paradis, 2005). At the early stages of development L2 children show evidence of transfer from their L1 to their L2 (Haznedar, 1997; Unsworth, 2005; Whong-Barr & Schwartz, 2002), but later on they seem to acquire the syntactic apparatus of their L2 (Haznedar, 2003).

To date, only two studies have focused on L2 children’s comprehension using off-line (Grueter, 2005) and on-line (Marinis, 2007) methods. These two studies have shown that despite L2 children’s problems with morphology in production, their off-line comprehension is target-like, and L2 children do not show qualitative differences from L1 children in the way they process morpho-syntax in real-time. The present paper goes one step further by investigating how L1 and L2 children process sentences involving reflexive and non-reflexive pronouns in real-time.
2. Acquisition and processing of (non)reflexive pronouns

From the 90s until now a considerable number of studies have investigated how children interpret sentences involving reflexive and non-reflexive pronouns. These have shown that monolingual children interpret reflexives in an adult-like fashion by the age of 3 to 4 years, but in some languages, such as English and Dutch, even 6 year-old children make errors in the interpretation of non-reflexive pronouns. E.g., in the Chien & Wexler (1990) study, 5 to 6 year-old children performed at chance level when they saw a picture of Mama Bear touching herself and were asked “Is Mama Bear touching her?”, showing that they allow the non-reflexive pronoun to be anaphorically linked to the local antecedent (Mama Bear).  

Several accounts have been proposed to explain this type of error. Grodzinsky & Reinhart (1993) have argued that these errors are caused by processing limitations. This is based on the idea that the interpretation of pronouns is regulated by Rule I, according to which if a message can be conveyed by two minimally different LFs, one involving variable binding and another one involving coreference, the variable binding structure is the preferred one. To apply this rule and reject a coreference interpretation, children have to hold two LFs in their working memory, which they sometimes fail to do due to processing limitations. Chien & Wexler (1990) and Avrutin & Wexler (1993) have taken a different standpoint arguing that children’s errors are caused by their difficulty to apply Principle P. Principle P is a pragmatic principle that prohibits coreference between two non-coindexed elements with the exception of elements appearing in Evans-style contexts. According to this view, children “know” Principle P, but they do not know the situations, in which the Evans-style reading is legitimate, which they will learn through experience.

The studies above have used off-line experiments to address the children’s comprehension of reflexive and non-reflexive pronouns. To date, only one study has investigated how children interpret reflexive and non-reflexive pronouns using off-line and on-line tasks (McKee, Nicol, & McDaniel, 1993). Using the on-line cross-modal picture priming paradigm and an off-line truth-judgment task, McKee et al. identified a group of 4 to 6 year-old children who showed a discrepancy between reflexive and non-reflexive pronouns in the off-line truth-judgment task and very poor performance in non-reflexive pronouns in both the match (50%) and mismatch conditions (18%). The same group of children showed a priming

---

1 For a review, see Guasti (2002), and for the discrepancy between production and comprehension see Hendriks (this volume).
effect for reflexives, but also a trend for a priming effect in non-reflexive pronouns with a local antecedent. This provides evidence that children who inappropriately accepted a local DP as the antecedent of non-reflexive pronouns in the off-line task, attempted to interpret a local DP as the antecedent of the non-reflexive pronoun also in the on-line task.

In adult L2 acquisition research, several studies have shown transfer effects for reflexive pronouns when the L1 grammar is the superset and the L2 the subset grammar (e.g., Hiragawa, 1990; Yip & Tang, 1998). For non-reflexive pronouns, White (1998) showed that finiteness affects the performance of Japanese learners of English at the intermediate level of proficiency, who reject local antecedents in biclausal finite contexts, but accept local antecedents in biclausal non-finite contexts. Finally, Guerel (2002) showed transfer effects in the interpretation of reflexive and non-reflexive pronouns in the subject position in English learners of Turkish.

The present paper investigates the comprehension of reflexive and non-reflexive pronouns in Turkish L2 children acquiring English and English controls using a task that combines on- and off-line measures. This can reveal the processes involved when we listen to sentences in real-time, and also the end product of comprehension in the same sentences.

3. Method

3.1 Participants

To address sentence processing in L1 and L2 children two groups of children were recruited, 28 6;10 to 8;8 year-old Turkish children with English as an L2 and 30 6;9 to 8;9 year old monolingual English children. Turkish-English children were selected from mainstream schools in North London and monolingual English children from schools in London and Sussex. We had as a selection criterion for the L2 children that both parents should be Turkish and Turkish should be the language spoken at home. The L2 children started systematic exposure to English when they went to nursery/school (at 3-to-4 years). Two standardised tests, the Test of Reception of Grammar 2 (TROG2) (Bishop, 2003) and the British Picture Vocabulary Scale (BPVS II) (Dunn, Dunn, Whetton, & Burley, 1997), were used in both groups of children as baseline tasks to ensure that the monolingual children had language abilities within the normal range, and also to assess the comprehension of grammar and single-word vocabulary of the L2 children. The BPVS II has an L1 and an L2 standardisation. Figure 1 shows z-scores for these tasks in the two groups.
Monolingual children performed in both baseline tasks within the normal range. Based on the L1 standard scores, L2 children scored more than 2 standard deviations (SD) below the mean in the comprehension of grammar and almost 2 SD below the mean in the comprehension of vocabulary. This is in line with previous research comparing language abilities of L2 children and children with SLI (Paradis, 2005). Based on the L2 standardisation, the L2 children performed within the norms.

### 3.2 Material

The experimental task combined picture-verification and self-paced listening, and was based on the task used in Marinis (2007). The children saw a picture on a laptop screen for 2500ms, and then listened to a sentence in a segment-by-segment fashion by pressing a button while the picture remained on the computer screen. Children were told that they have to press the button as quickly as possible in order to listen to the next segment. The end of each sentence was indicated by a beep sound. Prior to the experiment, the children heard a fixed set of instructions by the experimenter and then through the headphones. Then they were taught how to press the button to hear the sentences and were given 10 practice sentences to familiarise themselves with the task. The practice sentences could be repeated for a second time if the children or the experimenter felt that more practice was needed. E-prime was used for stimuli presentation and to record Reaction Times (RTs) from the onset of each segment until the button press.
The material comprised 10 practice, 20 experimental, and 40 filler sentences. The sentences were recorded by three female native speakers of English at a normal rate in the sound booth of our department, and were prepared for E-prime using Adobe Audition. Four different experimental sets were created, each containing one condition of each experimental sentence. Each participant encountered 5 sentences of each condition and was presented with only one of the four conditions of each item.

The experimental sentences included pronouns and reflexives. Agents and patients/themes were animals. 6 monosyllabic verbs were used two to four times throughout the experiment. Half of the sentences matched the picture. In the mismatch conditions for reflexives, the agent has performing an action on a different animal (e.g. a duck was tickling a bird). In the mismatch conditions of the non-reflexive pronouns, the agent was performing a reflective action (e.g., a duck was tickling himself). This makes a total of four experimental conditions (reflexive-match, reflexive-mismatch, pronoun-match, pronoun-mismatch), as shown below.

1. Condition 1: reflexive match, picture of a duck tickling himself
   I think / that / the duck / was tickling / himself / near the lake/ in July
2. Condition 2: reflexive mismatch, picture of a duck tickling a bird
   I think / that / the duck / was tickling / himself / near the lake/ in July
3. Condition 3: pronoun match, picture of a duck tickling a bird
   I think / that / the duck / was tickling / him / near the lake/ in July
4. Condition 4: pronoun mismatch, picture of a duck tickling himself
   I think / that / the duck / was tickling / him / near the lake/ in July

The sentences were ambiguous as to whether they match/mismatch the picture until the pronoun/reflexive, which was always in the fifth segment. The rationale underlying this task is that increased RTs to a particular segment indicate a processing difficulty at this point in the sentence. Mismatch between the picture and the sentence should cause a processing difficulty, and thus, increased RTs in Segment 5.

To ensure that the participants made an active effort to comprehend the stimulus sentences, at the end of each sentence children judged whether the sentence matched the picture. This was recorded by the tester on a separate form, and gave an additional off-line accuracy measure of comprehension. Children did not receive feedback as to whether or not they answered the comprehension question accurately. For half of the experimental sentences and fillers, the picture matched to the sentence and for the other half it did not. The experiment was carried out in the children’s school and lasted approximately 30 minutes.
On-line processing of pronouns in L1 and L2 children

4. Results

Prior to the analysis of the reaction times, the data were analysed for the accuracy in the off-line comprehension question. Table 2 shows the accuracy of the two groups in the 4 conditions.

<table>
<thead>
<tr>
<th>Conditions</th>
<th>L1 children</th>
<th>L2 children</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reflexive-match</td>
<td>95.7</td>
<td>89.1</td>
</tr>
<tr>
<td>Reflexive-mismatch</td>
<td>94.1</td>
<td>86.8</td>
</tr>
<tr>
<td>Pronoun-match</td>
<td>42.3</td>
<td>67.1</td>
</tr>
<tr>
<td>Pronoun mismatch</td>
<td>27.5</td>
<td>25.2</td>
</tr>
</tbody>
</table>

Table 2. Off-line accuracy data in percentage

A repeated measures ANOVA with the factors Group (L1, L2), Sentence Type (reflexive, pronoun) and Matching (match, mismatch) showed a main effect of Sentence Type (F (1, 56) = 141.611, p < .001) reflecting better accuracy in sentences involving reflexives compared to pronouns, and a main effect of Matching (F (1, 56) = 33.961, p < .001) reflecting better performance in matched compared to mismatched sentences. Interactions between Group x Sentence Type (F (1, 56) = 4.542, p < .05), Group x Matching (F (1, 56) = 7.169, p = .01), Sentence Type x Matching (F (1, 56) = 68.496, p < .001), and Group x Sentence Type x Matching (F (1, 56) = 16.963, p < .001) showed that the two groups performed in a different way in the four conditions. Tests of simple effects showed that for sentences with reflexives both groups performed equally well in the match and mismatch conditions (p > .1), but they performed better for sentences with pronouns in the matching compared to the mismatch condition (p < .001). In addition, both groups performed better in sentences involving reflexives compared to sentences involving pronouns in both the matching and the mismatch condition (p ≤ .001). However, there were also important differences between the two groups, which caused the three-way interaction. L2 children performed significantly worse than L1 children in the matching condition for reflexives (p < .05), but they performed significantly better than L1 children in the matching condition for pronouns (p = .01).

RTs for accurate responses were analysed separately for each segment in repeated measures ANOVAs with the factors Group (L1, L2), Sentence Type (reflexive, pronoun) and Matching (match, mismatch). Due to space limitations, this paper will focus only on the critical segment (Segments 5) and the following segment (Segment 6) that can show a spill-over effect.
Figure 2 shows the mean RTs for accurate responses of the L1 children and Figure 3 the mean RTs per for accurate responses of the L2 children.

In Segment 5, the repeated measures ANOVA showed a main effect of Sentence Type (F (1, 56) = 42.617, p < .0001) reflecting overall longer RTs for the reflexives, a main effect of Matching (F (1, 56) = 13.245, p = .001) reflecting overall longer RTs in the Mismatch compared to the Match condition, and a significant interaction between Group x Sentence Type (F (1, 56) = 19.494, p < .0001) reflecting differences between the two groups in the RTs for reflexives vs. pronouns. Analyses of simple
effects revealed shorter RTs for reflexives in L1 compared to L2 children (p < .05), but there was no significant difference between the two groups for pronouns (p > .1). In addition, L2 children showed longer RTs for reflexives compared to pronouns (p < .0001), but there was no significant difference between reflexives and pronouns for L1 children (p > .1).

In Segment 5, the repeated measures ANOVA showed a main effect of Group (F (1, 56) = 4.934, p < .05) reflecting overall longer RTs for L2 compared to L1 children, a main effect of Sentence Type (F (1, 56) = 4.927, p < .05) reflecting overall longer RTs for the reflexives compared to pronouns, a main effect of Matching (F (1, 56) = 8.589, p = .01) reflecting overall longer RTs in the Mismatch compared to the Match condition, and a 3-way interaction between Group x Sentence Type x Matching (F (1, 56) = 4.250, p < .05) reflecting differences between the two groups in the RTs for reflexives vs. pronouns. Analyses of simple effects revealed a spill-over effect in L1 children, who showed longer RTs in the mismatch compared to the matching condition for pronouns (p < .01), but not for reflexives (p > .1). L2 children showed no spill-over effect for either of the two (p > .1). L1 children also showed longer RTs for pronouns in the mismatch condition compared to reflexives in the mismatch condition (p < .01) reflecting processing difficulties when there was a mismatch between the picture and the sentence for pronouns.

5. Discussion

The results from the baseline tasks revealed that L2 children performed at a significantly lower level than the average monolingual children in both the comprehension of grammar and vocabulary, but their vocabulary scores were age equivalent for their L2 status.

The off-line accuracy scores from the sentence processing task replicated previous findings from L1 studies using off-line tasks. L1 children performed at ceiling in reflexive, but poorly in non-reflexive pronouns. Poor comprehension on non-reflexive pronouns was attested not only in the mismatch (28%), but also in the matching condition (42%), a similar pattern to the study by McKee, et al. (1993). The same overall pattern was attested also in the L2 children, but L2 children performed significantly better than the L1 children in the non-reflexive matching condition (L2: 67%, L1: 42%). This is in the first instance surprising given the L2 children’s low score on the comprehension of grammar in contrast to the L1 group, who performed within norms. How can this difference be explained? Better performance of the L2 children could relate to the properties of pronouns in their L1. According to Guerel (2002), Turkish
has overt and null pronouns. Overt pronouns in the object position have the same binding properties as non-reflexive pronouns in English, i.e. they cannot be bound by a local antecedent. In contrast, null pronouns can have a local, or a non-local antecedent. Turkish children may transfer the properties of the Turkish overt pronouns to the English non-reflexive pronouns, and this could account for their higher accuracy in non-reflexive pronouns in the matching condition.

In terms of RTs, when the children were accurate in the interpretation of reflexive and non-reflexive pronouns, both groups showed longer RTs for both types of pronouns in the critical segment of the mismatch compared to the matching condition. This shows no qualitative differences between the two groups despite the lower scores of the L2 children in standardised assessments of grammar and vocabulary. Thus, both groups seem to process and interpret reflexive and non-reflexive pronouns rapidly. This finding is in line with previous findings on L2 children’s on-line processing (Marinis, 2007) showing that they process sentences in a similar way to L1 children.

Acknowledgments

This study was supported by the Research Endowment Trust Fund from the University of Reading (E3097100). Many thanks to Christos Pliatsikas and Emine Topal for the data collection, and the children and schools for their participation.

References

Chien, Y.-C., and K. Wexler (1990) “Children's knowledge of locality conditions in binding as evidence for the modularity of syntax and pragmatics,” Language Acquisition 1, 225-295.


