Variation, Asymmetry and Working Memory in the Process of Second Language Acquisition

Clare Wright
Newcastle University, UK

1. Introduction

Calls for understanding the interface between L2 linguistic knowledge and development (Gregg 1996; Carroll 2001; Towell 2003) provide a context for analysing the role of memory (Paradis 2004), specifically working memory (Baddeley 1986, 2003) in L2 development. Miyake and Friedman (1998) have claimed that Working Memory (WM) may be the key to L2 acquisition, especially in explaining individual variation in L2 acquisition. Recent findings found a robust connection between greater working memory (WM) capacity and rapid, successful acquisition of L2 vocabulary, reading and oral fluency (Service 1992; Harrington and Sawyer 1992; Fortkamp 1999). This study adds to the growing body of research by investigating correlations between WM and variation in grammatical development, focusing on asymmetries in processing L2 English wh-constructions in an immersion setting.

2. Theoretical and Empirical Background

English wh-question formation has been identified as a source of wide variation in L2 acquisition, particularly for Chinese speakers of English (CSE), even at advanced levels and after immersion in English (Schachter & Yip 1990; Johnson & Newport 1991). This variation may be due to factors such as L1 transfer or general developmental constraints. Wh-questions require syntactic movement while Chinese lacks overt wh-movement; acquisition of the L2 linguistic knowledge to produce accurate wh-questions is seen as late acquired (Pienemann 1998), especially for long-distance movement such as What did John say Mary wanted? or constrained by subjacency, such as *What did the book about please Mary? However, factors affecting inter-learner variation are less well understood (Han 2004; Wright, 2006, 2009).

Standard generative accounts of whether CSE acquire target-like wh-movement remain inconclusive (see, e.g., White & Genesee 1996; Hawkins & Chan 1997; White 2003; Schwartz et al. 2008). Other accounts of wh-movement suggest that individual differences in general cognitive processing explain individual variation (Johnson & Newport 1991; Clahsen & Felser 2006; McDonald 2006). The question of how input triggers acquisition of the L2 also remains debated (Schwartz 1993; Truscott & Sharwood Smith 2004; Carroll
CSE have commonly been taught in an input-poor learning environment, where a traditional emphasis on explicit and written linguistic knowledge can result in wide variation between oral and written proficiency (Ellis 1994; Gu 2003). However, research suggests that CSE can reach native-like competence even without exposure to native language immersion (White & Juffs 1998). Focusing on variability in acquisition of wh-movement in an immersion setting should therefore provide insight into how instructed learners may vary in processing linguistic knowledge and how such processing may change over time as a result of differences in exposure provided within the context of an immersion setting.

The aim of this study is to examine this assumption by comparing L2 learners’ accuracy on explicitly taught grammatical structures (short-distance wh-movement, grammatical long-distance wh-movement), compared to ungrammatical structures that cannot be taught and are argued to be rare in the input (subjacency-constrained movement). For instructed learners, L2 linguistic knowledge may be accessible via a coalition of resources (Herschensohn 2000). Although the psychological constructs underpinning such a coalition are complex, in simplified terms instructed learners may draw either on linguistic competence (implicit procedural memory), or learned linguistic knowledge (explicit declarative memory), as suggested by, amongst others, Schwartz (1993), Ullman (2001), Paradis (2004).

The inference drawn here is that L2 users with little naturalistic L2 input, but using ‘instructional bootstrapping’ (Herschensohn 2000: 220), will initially store L2 linguistic knowledge of morphology and syntax primarily as consciously accessible or explicit knowledge. After sufficient exposure to primary linguistic data, implicit knowledge develops, subject to the UG template. Until the L2 user can utilize the quicker, more efficient but non-accessible implicit system, it is assumed that conscious access of explicit knowledge is key for the L2 user. If conscious access is required, then WM, the temporary ‘workspace’ for conscious attention to complex tasks, will also be key in efficient retrieving or inhibiting existing explicit knowledge and processing novel information (Smith & Kosslyn 2007: 247).

Much of the research on WM in native language and L2 acquisition has been based on versions of Baddeley and Hitch’s (1974) multi-component model. The latest model (Baddeley, 2000, 2003) posits domain-specific temporary storage via the phonological loop and the visuo-spatial sketch-pad. Domain-general attentional control and processing efficiency is managed by a central executive, with an episodic buffer allowing domain-general storage for more than the standard 1-2 seconds (see Figure 1 below). The episodic buffer, a new element (Baddeley 2000), is designed to explain how novel and retrieved information can be combined and maintained, for example allowing a prose passage of around 90 seconds to be retained and repeated accurately. WM is seen as capacity-constrained: as storage capacity reaches or exceeds its limit, processing efficiency is reduced, so greater storage capacity allows greater processing efficiency.
Figure 1: Baddeley’s (2000) Multicomponent Model of Working Memory

Initial research findings using WM tests for phonological loop storage and central executive efficiency have found a robust correlation between WM and native-language (L1) proficiency (Daneman & Carpenter 1980; Gathercole & Baddeley 1993; Gathercole 2006). This robustness extends to certain aspects of L2 proficiency: for vocabulary acquisition, reading comprehension, resolving syntactic ambiguities and oral fluency (Harrington & Sawyer 1992; Service 1992; Miyake et al 1994; Ellis & Sinclair 1996; Baddeley et al. 1998; Miyake & Friedman 1998; Fortkamp 1999).

However, the role of WM in L2 grammatical development has not been widely studied and remains unclear (Sagarra 2000; Juffs 2004). There is also some debate over Baddeley’s multi-component model used in the L2 studies cited above (Andrade 2001; Caplan & Waters 1999; Cowan 1999; Miyake & Shah 1999). Nevertheless, Baddeley’s model remains the most widely used for researching WM in language acquisition and is thus the basis for this study, which specifically addresses the question of identifying a suitable methodology for testing the episodic buffer. Investigating the role of WM in grammatical development, as well as the need to developing appropriate testing methodologies, are thus the two goals driving this study.

Current evidence leads to two assumptions underpinning the research presented here as to how WM may be key to L2 acquisition even at an advanced level. The first relates to the role of WM in attentional control, where WM acts as a kind of ‘bottle-neck’ (Emerson et al. 1999) through which L2 linguistic operations have to pass – the more novel the sound, or the more complex the task, the more significant the capacity of the ‘bottle-neck’. For L2 users, processing complex morpho-syntax under pressure, such as in timed grammaticality tasks, requires conscious control over accessing explicit L2 knowledge and inhibiting L1 language patterns. The second assumption is that the more difficult the morpho-syntax (such as for long-distance wh-movement), the greater the effort in processing accurate forms. This assumption is supported by research into the role of WM in native-language complex syntax, such as assigning relative-clause reference (Miyake et al. 1994) and subordination and adverbial use (Fry 2002; Fehringer and Fry 2007). Furthermore, processing
difficulties are hypothesized as an explanation for individual variation in L2 (Cook 1997; Service et al. 2002). This processing account is found even in generative studies such as White and Juffs (1998), who explained evidence of variation in native-like L2 acquisition as ‘implicit competence processed more slowly’ (ibid: 127).

English question forms provide a suitable template for studying WM in L2 acquisition in terms of how WM may facilitate access different types of linguistic knowledge, particularly the asymmetry highlighted by Schwartz (1993), and Ullman (2001) between explicit or declarative knowledge and implicit or procedural knowledge. Question forms are explicitly taught in Chinese schools from junior high school (Nani 2006); short movement is presented earliest and most frequently; long movement questions are more rarely presented (ibid). Violations of subjacency constraints are by definition untaught and evidence to trigger awareness of such constraints argued to be rare in the input (White 2003). WM is argued (Baddeley 2000) to prioritise manipulation of explicit or declarative knowledge, and can therefore be claimed to facilitate asymmetric development between taught knowledge (such as short movement) and untaught knowledge. WM capacity can be expected to play a role in aiding improvements in processing explicitly taught knowledge, particularly of short-distance movement, but would not be expected to be associated with processing of untaught subjacency constraints.

3. Study Design and Participants

To investigate the conceptual and methodological issues raised above, a semi-longitudinal study was designed to test for positive correlations between WM and L2 variation in acquisition by adult Chinese speakers of English immersed in UK while undertaking a year’s postgraduate study at British universities. Two research hypotheses were addressed:

(i) Immersion facilitates acquisition of wh-movement, subject to asymmetries between taught and untaught forms;

(ii) WM correlates with greater accuracy in processing wh-movement in an immersion environment, but only for taught forms, not untaught forms.

Thirty-two advanced adult speakers of English, with Mandarin Chinese as L1, were recruited from a cohort of newly arrived postgraduates at UK universities. All were instructed learners with no previous immersion exposure, with a minimum IELTS score of 5.5 in the previous four months. The group consisted of eighteen participants from Taiwan, and fourteen from Mainland China; there were eight males and twenty-four females. Mean age of starting learning English was 11.41 years old (SD 1.58) and mean length of exposure was 11.77 years (SD 2.91). Linguistic data was collected at Time 1 (within two weeks of arrival) and at Time 2 (after ten to eleven months’ immersion). WM data was collected at Time 1 and at Time 2. Since a directional correlation was
assumed between WM and linguistic development, only WM scores from Time 1 were used in the analysis reported here (following Sagarra 2000).

Linguistic data were collected using a self-paced computerised timed grammaticality judgement task, adapted from White and Juffs (1998) and Wright (2009). Participants were asked to indicate the grammatical acceptability of 68 question forms testing three types of wh-question. There were 40 tokens testing short movement (e.g. *What did John eat?), consisting of 20 grammatical and 20 ungrammatical tokens. There were 16 tokens testing grammatical long movement (e.g. What did you think Mary saw?) and 12 tokens testing ungrammatical subjacency violations (e.g. *What did the book about please Ann?). There were also 14 distractor items. The task was created using DMDX software, version 3.1.6.2 (Forster and Forster 2003). Participants’ responses were encoded as they pressed specific computer keys to make their judgements on randomly presented items, which measured the speed and targetlike accuracy for each item.

For WM data collection, in line with current best practice in testing WM (see, e.g., Conway et al. 2005), a battery of tasks was used: a non-verbal task (Digits Back), and two innovative verbal tasks designed for this study (Story Recall, and an adapted Listening Span task). The Listening Span task proved to be unreliable in showing individual differences, so only Digits Back and Story Recall scores are reported here. Both tasks were conducted in Mandarin and English in order to shed further light into investigations as to how far WM is language-independent (Osaka & Osaka 1992; Chincotta and Underwood 1997) or affected by differences in L2 processing (Cook 1997; Service et al. 2002).

The first task, Digits Back, was chosen as being widely used (Waters & Caplan 2003), and therefore acted as a reliable benchmark, easy to administer, and unrelated to linguistic proficiency (since advanced learners would all be familiar with the English digit names). Participants heard sets of numbers, increasing in length from four to seven, read out at a rate of one digit per second (two strings per set) first in English and later in Mandarin. After each string, participants repeated the numbers in reverse order. Scoring was calculated using partial credit scoring, as a ratio of correct responses between 0 and 1 (Conway et al. 2005: 774).

The Story Recall tasks were created to address the issue of how to test newer models of the WM construct, in particular Baddeley’s episodic buffer, for which virtually no research on L2 WM has yet been published. The Story Recall tasks devised for this study were based on standard psychology tests (see, e.g., Coughlan & Hollows 1985), which measures the accuracy in recalling prose passages of over 30 seconds in length. This test had been identified as correlating with use of complex syntax such as subordinate clauses and adverbial phrases in native language use (Fry 2002) and bilingual language use (Fehringer and Fry 2007). The purpose of using the task in this study was to

---

1 DMDX software, developed by Ken and Jonathan Forster at Monash University and University of Arizona, is available at http://www.u.arizona.edu/~jforster/dmdx.htm
assess whether it was also a reliable, valid means of testing WM in less proficient L2 users, and if WM as measured by such a test correlated with the use of complex question formation. The original Coughlan and Hollows test was adapted and translated into Mandarin (lasting 54 seconds). A different story in English (with similar schematic structure) was devised by the researcher (lasting 33 seconds). The length of the L2 version was shortened to avoid possible ‘floor effects’ due to task difficulty (Harrington & Sawyer 1992: 28). Two bilingual raters worked with the researcher in scoring the Mandarin data to ensure scoring reliability. Scoring for the task was out of 50 for accurate recall of morpho-syntactic and semantic elements.

4. Results and discussion

All the scores for reaction-time data and WM data were encoded using SPSS. Non-parametric statistical tests were used, due to non-normal distribution found on the majority of the task scores. Tables 1 and 2 below show the mean, minimum and maximum scores for overall accuracy and speed on the reaction-time task at Time 1 and Time 2. Reaction time speeds are shown here in seconds rather than milliseconds, for ease of reference.

Table 1: Mean overall accuracy

<table>
<thead>
<tr>
<th></th>
<th>Accuracy/68 (%)</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>40.13 (59.01%)</td>
<td>10.76</td>
<td>20</td>
<td>57</td>
</tr>
<tr>
<td>Time 2</td>
<td>39.69 (58.36%)</td>
<td>8.63</td>
<td>18</td>
<td>56</td>
</tr>
</tbody>
</table>

Table 2: Mean overall speed

<table>
<thead>
<tr>
<th></th>
<th>RT (in secs)</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>506.37</td>
<td>159.17</td>
<td>155.04</td>
<td>846.19</td>
</tr>
<tr>
<td>Time 2</td>
<td>432.81*</td>
<td>151.75</td>
<td>235.90</td>
<td>997.02</td>
</tr>
</tbody>
</table>

*significantly different to Time 1 (p<.05)

As outlined above, three different types of wh-question were under investigation (short movement, long movement and subjacency violations). In order to compare the three specific types, which were not equally balanced for number of tokens, RT scores on speed and accuracy by type were recalculated to give an average per item of each type. The average item scores showing accuracy and speed by type at Time 1 and Time 2 are given in Tables 3 and 4 below.

Table 3: Average speed (in secs) by type at Time 1 and Time 2

<table>
<thead>
<tr>
<th></th>
<th>Short mot v</th>
<th>Long mot</th>
<th>Subjacency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>5.993</td>
<td>7.222</td>
<td>7.768</td>
</tr>
<tr>
<td>Time 2</td>
<td>5.1</td>
<td>5.932</td>
<td>6.758</td>
</tr>
</tbody>
</table>
Table 4: Average accuracy (%) by type at Time 1 and Time 2

<table>
<thead>
<tr>
<th></th>
<th>Short movt</th>
<th>Long movt</th>
<th>Subjacency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>57.42%</td>
<td>46.88%</td>
<td>44.79%</td>
</tr>
<tr>
<td>Time 2</td>
<td>58.67%</td>
<td>51.17%</td>
<td>35.94%</td>
</tr>
</tbody>
</table>

These data confirmed the asymmetries predicted in Hypothesis 1, that short movement would be judged more accurately and faster than other types at both Time 1 and Time 2, but these asymmetries were nonsignificant (p>.05). However, against expectation, improvement was only found to any marked degree for Long movement, and accuracy in subjacency in fact decreased by Time 2; again, these changes were nonsignificant (p>.05). In sum, the expected linguistic development resulting from immersion was not found.

Turning to investigate the role of WM predicted in Hypothesis 2, the lack of significant change in accuracy on the RT task could be expected to confound the assumptions underpinning Hypothesis 2. The expectation was that linguistic data were expected to show significant correlations between greater WM capacity as measured by Digits Back and Story Recall, but only for taught forms, not for subjacency violations, both at Time 1 and Time 2. It was also expected that correlations would be similar for both L1 and L2 versions of the WM tests, showing that WM is language-independent.

The Digits Back tasks in either L1 or L2 showed no significant correlations for either RT speed or accuracy (p>.1), and are therefore not reported here. Story Recall (SR) scores in both L1 and L2, on the other hand, were found to correlate significantly with the linguistic data. Significant correlations were found distributed across all three types of wh-questions at both Time 1 and Time 2, as shown in Tables 5 to 8 below.

Table 5: WM correlations with accuracy, Time 1

<table>
<thead>
<tr>
<th></th>
<th>Overall accuracy</th>
<th>Short movt</th>
<th>Long movt</th>
<th>Subjacency</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR L1</td>
<td>0.192</td>
<td>0.247</td>
<td>-0.439*</td>
<td>0.402*</td>
</tr>
<tr>
<td>SR L2</td>
<td>-0.032</td>
<td>-0.05</td>
<td>-0.155</td>
<td>0.014</td>
</tr>
</tbody>
</table>

*Correlation is significant (p<.05)

Table 6: WM correlations with accuracy, Time 2

<table>
<thead>
<tr>
<th></th>
<th>Overall accuracy</th>
<th>Short movt</th>
<th>Long movt</th>
<th>Subjacency</th>
</tr>
</thead>
<tbody>
<tr>
<td>SR L1</td>
<td>-0.043</td>
<td>0.116</td>
<td>-0.437*</td>
<td>-0.128</td>
</tr>
<tr>
<td>SR L2</td>
<td>0.017</td>
<td>-0.145</td>
<td>-0.128</td>
<td>0.277</td>
</tr>
</tbody>
</table>

*Correlation is significant (p<.05)
As shown in these tables, there was a mixture of both positive and negative correlations found between Story Recall scores and the RT scores, distributed across each wh-question type. None of the expected associations were found on RT accuracy; there was no correlation found with Short movement, and a significant negative correlation found with Long movement at both Time 1 and Time 2. The only significant positive correlation for accuracy was found with Subjacency scores at Time 1, counter to prediction. The most consistent pattern was of positive correlations between RT speeds (at both Time 1 and Time 2) and Story Recall scores, shown in Tables 7 and 8 above. Correlations between Story Recall and slower RT speeds were mainly associated with taught forms at Time 1, but showed similar patterns across all forms by Time 2. In other words, higher Story Recall scores correlated with slower RT speeds regardless of type of wh-question (taught or untaught); again, this was counter to predictions. It also seems clear that WM is language-independent (Osaka & Osaka 1992), since Story Recall correlations were very similar, whether for L1 or L2. In addition Story Recall scores correlated significantly cross-linguistically at Time 1 ($r=.402$, $p<.05$), as did Digits Back ($r=.569$, $p<.001$).

The correlations shown above disconfirmed the predictions of Hypothesis 2, since WM was not found to facilitate processing only of explicitly taught forms. The only significant positive correlation between WM and RT accuracy was found on subjacency scores; for RT speeds, correlations were only found on between WM and slower times, and this pattern was found across all types of question forms. It is thus concluded that WM, as measured by the Story Recall tasks described here, may be argued to operate as a general “workspace” for manipulating both explicitly taught and untaught implicit knowledge, rather than only favouring explicit knowledge.

These findings present a challenge to some of the assumptions drawn in this study based on psychological models of learning and memory and the role of WM in linguistic development and processing (Jackendoff 1997; Ellis 2005). The assumed link between taught input and explicit or declarative memory storage, inferred in Baddeley’s (2000) model of WM, may not be as specific as
claimed here. In addition, the reliance on declarative memory by non-proficient L2 learners argued by Ullman (2001) may also not be as robust as claimed. Heredia and McLaughlin (1992) identified that the typical psycholinguistic separation between explicit (or declarative) and implicit (or procedural) linguistic representations may be difficult to identify or reliably testable. The evidence from this study seems to indicate that WM can aid the development of linguistic knowledge, but that the concept of WM as a workspace involving conscious “focused attention” (Cowan 1999) needs to be able to incorporate both implicit linguistic knowledge (e.g. subjacency violations) and explicit knowledge (e.g. short-distance question formation rules).

In methodological terms, the lack of Digits Back to show any relation with the linguistic tasks may be due to WM being domain-specific (as argued by Caplan and Waters 1999). However, the robust association of Story Recall with reaction times presents a new metric to investigate further to see how this innovative task can support future research into the nature of WM (Fehringer and Fry 2007).

However, the most interesting finding from this study was the lack of significant development in accuracy on the reaction time task during a year’s immersion, suggesting also that further research is needed into the nature of linguistic development, and how far the type of immersion provided within a typical postgraduate environment provides adequate input to trigger acquisition.

5. Conclusion

This study was an innovative investigation into the nature of representation and processing in L2 acquisition and the interface between morpho-syntactic development and working memory (WM). A reaction-time grammaticality judgement task was used to test for effects of a year’s immersion in the UK. Results showed predictable patterns of variation in acquisition of English wh-questions, in line with a hypothesised asymmetry between explicitly-taught and frequently presented question forms (e.g. short-movement) and untaught implicitly-acquired forms (subjacency violations). Overall, mean linguistic development was observed only in terms of faster processing times, rather than increases in accuracy. This study concludes that immersion appears to favour processing existing grammatical knowledge more efficiently rather than trigger acquisition of new grammatical knowledge, even for those with greater WM. The study also found, counter to expectation, that WM correlated significantly (p<.01) with longer reaction times across all forms, and with greater accuracy on untaught forms (subjacency violations). These results support Herschensohn’s (2000) claim that L2 users depend on “a coalition of resources” including implicit UG-constrained knowledge and explicit instructional bootstrapping. The findings also indicate that further research is required to understand the nature of linguistic development, particularly the complex interface between input and memory, and how precisely WM is involved in the process of second language acquisition.
References


