Are All L2 Learners Morphologically Insensitive? L1 Influence and WM Capacity as Mitigating Factors

Guillermo Rodríguez
University of Vermont

1. Introduction

Second language (L2) sentence processing has been claimed to be shallower than native sentence parsing (Clahsen & Felser, 2006; but see Witzel, Witzel & Nicol, 2012 for evidence to the contrary). In other words, L2 processing is said to lack the configurational detail based on syntactic information that is characteristic of L1 parsing. Additionally, L2 parsing has also been described as less accurate due to factors such as proficiency level, individual cognitive differences and speed of lexical access (McDonald, 2006). Recent findings have also suggested that L2 learners are morphologically insensitive, not being able to detect word form inaccuracies while processing sentences online (Foote, 2011; Jiang, 2004; 2007; Keating, 2009; Sagarra & Herschensohn, 2010). If generalizable, this lack of sensitivity to morphology would add another variable to consider in the study of L2 sentence processing (Hopp, 2006; 2010). This paper investigates the morphological sensitivity of English-speaking learners of Spanish, following Jiang's (2007) experimental design, with the addition of language-independent working memory (WM) measures.

2. Literature review

Automatic or integrated knowledge in an L2 has been described as the linguistic knowledge that learners can retrieve and apply without deliberate effort (DeKeyser, 1997; Robinson, 1997; Ullman, 2001). Within this framework, L2 learners have been characterized as unable to integrate (or automatize) morphological information, since they do not seem to notice mistakes dealing with subject/verb agreement, and gender/number markers. However, they do detect argument structure errors during self-paced reading (Jiang, 2004; 2007).

In Jiang's (2007) study, 26 Chinese-speaking learners of English read 64 sentences involving grammatical errors with the plural –s marker (1) and verb subcategorization mistakes (2) following a self-paced moving window paradigm, which was implemented in order to minimize the involvement of explicit knowledge (Jiang, 2004). The performance of the L2 learners was compared to that of an equal number of native speakers (NS) of English.

(1) * The visitor took several of the *rare* *coin* *in* *the* cabinet.
(2) * The teacher insisted the *student* *to* *start* *all* over again.

The underlined words in (1) and (2) show the regions of interest, with the second word being the region in which either a morphological error – *coin* in (1) - or a subcategorization mistake – *to* in (2) - was located. Jiang recorded reaction times (RTs) for both the preceding region and the two regions following the target, in order to assess any spillover effect after the error loci. Each ungrammatical sentence was counterbalanced with a grammatical version and no participant saw the two versions of a particular item. Half of the stimuli presented were followed by comprehension questions to ensure that participants were reading for meaning. The analyses of RTs showed that NSs were sensitive to both types of errors, and this sensitivity translated into longer RTs on the third and fourth regions of interest, not on the actual word that bore the error. On the other hand, the Chinese speakers showed no significant differences between the grammatical and ungrammatical stimuli for sentences like (1) with morphological errors, but did show a significant difference for items like (2), also in the spillover region. These findings led Jiang to claim that L2 learners are not sensitive to the lack of required morphological markings in their target language; in other words, they have not integrated morphological information as an automatic component of their L2 sentence comprehension repertoire. Other recent studies have looked at the relationship between morphological sensitivity and other variables that have to do with the stimuli used and the proficiency of the participants involved. For example, Keating (2009) used eye-tracking to test sensitivity to gender agreement morphology among L2 learners of Spanish when reading sentences online. The distinguishing feature of this study was the addition of distance between the agreement elements. The stimuli in Keating's task included grammatical and ungrammatical agreement relationships between a noun and an adjective that were adjacent, as in (3), or not, as shown in (4).

(3) * Una fiesta pequeño es ideal para una persona tímida o introvertida.
A party(f) small (m) is ideal for a shy or introverted person.
(4) * Una casa es bastante pequeño cuando tiene sólo una habitación.
A house(f) is rather small(m) when it only has one bedroom.

Both variables, proficiency and distance, proved significant in Keating´s results. Although learners with high proficiency performed similarly to native speakers, it was only when the agreement relationship involved immediately adjacent elements.

While trying to tease apart the relationship between gender and number agreement errors, Sagarra & Herschensohn (2010) tested the sensitivity of L2 learners of Spanish to number and gender agreement with the word-by-word self-paced reading paradigm. The agreement discrepancies in this study involved contiguous adjectives and nouns within a noun phrase in direct object position (as shown in (5)).

(5) El ingeniero presenta el prototipo famoso/*famosa//famoso/*famosos en la conferencia.
The engineer shows the famous prototype at the conference.

Contrary to what Jiang (2007) and Keating (2009) found, Sagarra & Herschensohn's findings showed that intermediate learners of Spanish were sensitive to agreement violations of both types—gender and number—in an adjacent relationship.

In a similar study, Foote (2011) used the self-paced moving window paradigm to test advanced learners of Spanish, who read sentences involving gender agreement mismatches in local (6) and non-local dependencies (7).

(6) Dicen que el libro blanco/*blanca está en esa mesa.
They say that the white book is on the table.
(7) El pollo del taco está rico/*rica pero picante.
The chicken in the taco is tasty but spicy.

The learners in this study showed sensitivity to agreement mistakes in both the local and the non-local stimuli. Foote's findings suggest that native-like processing of this type of morphology is possible in advanced learners.

The studies reviewed above show contrasting results as to the capacity of L2 learners to represent and process morphological agreement in a native-like fashion in the target language with both local and non-local dependencies. However, the lack of a consistent pattern allows for further investigation of the issue with a different agreement relationship, for example, person agreement between the subject and the main verb.

Additionally, individual differences that may affect the deployment of morphosyntactic knowledge, such as WM capacity (McDonald, 2006), may help untangle the interaction between proficiency and stimuli types that has been attested in previous research.

WM capacity has been claimed to aid in the acquisition of vocabulary and the comprehension of sentences in an L2 (see Gathercole, 2007 for a comprehensive review). However, no studies to date have tested its influence in the integration of person verbal morphology information during parsing.

Within the multi-component model of WM (Baddeley, 2003), the central executive is supposed to control the flow of memory resources among tasks, and it is understood to be a limited-capacity component. If selective integration of different grammatical components is a gradual or developmental process in the L2, as claimed by Jiang (2007), we would expect that knowledge that has not been automatized would require WM resources for its processing. Thus, an individual with higher WM resources could devote more of that capacity to not only try to integrate the syntactic requirements for a predicate (involved in subcategorization mistakes), but also to pay attention to form, and particularly to agreement relationships between words (the morphological errors involved in the present study). In this respect, an intermediate L2 learner whose developing proficiency had not yet integrated morphological knowledge would benefit from a higher WM capacity in order to be able to process word form.

One of the preferred instruments to test the capacity of the central executive is the Reading Span Task (RST), developed by Daneman & Carpenter (1980). In this task, participants read sentences for meaning and are asked to maintain in memory either the last word of each sentence read or an unrelated word. However, this task has been problematic, since the pace is usually controlled by the participant (allowing for the use of strategies). Moreover, it is language dependent (meaning that language mediates the results of the test; see Engle, Nations, & Cantor, 1990), and the task is administered in the L2 learners' first language (L1). These issues prompted the use of a different instrument to assess WM capacity in the L2 learners in the current study: the Automated Operation Span Task (AOST, Unsworth, Heitz, Schrock & Engle, 2005). This test is language independent, it is timed without participant control, and it involves processing and storage, both of which are measures of the central executive (Conway, Kane, Bunting, Hambrick, Wilhelm & Engle, 2005). As mentioned above, the motivation for the adoption of this test in the present study was to avoid the pitfalls of the traditional RST task. However, the AOST has only been used once before in the assessment of WM for L2 studies (Rai, Loschky, Harris, Peck & Cook, 2010); thus, the variety of instruments used makes generalizations difficult to interpret in the L2 literature on WM (Juffs & Harrington, 2010). Therefore, in order to make sure that the task chosen was actually measuring the purported processing and storage capabilities of the central executive, the Raven's Advanced Progressive Matrices Test (RPMT; Raven, Raven & Court, 1998) was administered to participants in addition to the AOST. The former is a measure of abstract reasoning (or fluid intelligence) that consistently correlates with WM capacity with averages of r=.30 (Wiley, Jarosz, Cushen & Colflesh, 2011). Consequently, if a similar relationship between the AOST and the RPMT was found in this study, it would increase the validity of the instrument chosen here.

3. Study motivation and research questions.

This paper partially replicated Jiang's study (2007), with a different L2 group and the application of a psychometric approach to test WM capacity and fluid intelligence. The empirical tests included here were meant to answer the following research questions:

R1. Is morphological insensitivity dependent on language background? In other words,, does the presence of similar morphological markings between the L1 and the L2 of the learner ameliorate their performance when processing morphology in the target language?

R2. Is person verbal morphology also a category subject to insensitivity in L2 processing?

R3. Is an individual with high WM capacity better able to process morphology in their L2?

In order to answer these questions, eighteen English-speaking learners of Spanish took a word-by-word self-paced reading comprehension test meant to measure their reading times (RTs) for sentences involving morphological mistakes, as in (8):

(8) *La perra duerm***o** en el garaje cuando hace frío.
**The dog sleep in the garage when it's cold.*

According to previous findings (Jiang, 2004; Pearlmutter, Garnsey & Bock, 1999), readers are expected to take longer to read regions on or immediately following the onset of the error (the spillover effect), if detected in the input. In this case *duermo*, a verb inflected for first person singular and simple present tense, does not agree with the third person subject. When compared to the same sentence without an inflectional mistake, this error should trigger longer RTs on the actual verb or the following regions/words.

The inclusion of a different group of L2 learners was motivated by the need to test a morphological marking that is present in both the native language and the target language of the participants tested, in order to discount issues of cross-linguistic influence. Chinese seldom makes use of plural marking (Sun, 2006); thus, an almost missing category in the L1 may have increased the difficulty of its processing or integration in the L2. On the other hand, both English and Spanish require some kind of morphological agreement between the subject and the main verb of a clause. This relationship, just as plural marking for English learners (Jiang, 2007; Keating, 2009), is something that is emphasized and practiced from the beginning levels of instruction in Spanish as a foreign language. As already mentioned, this replication also included both a test of WM capacity and a test of fluid intelligence, which were meant to characterize the nature of the relationship between WM resources and the selective integration of morphological knowledge in an L2.

4. Methods

4.1. Participants

A total of 25 English-speaking learners of Spanish were tested in the study. All were students in two sections of an intermediate Spanish class at a public university in the US. Their Spanish proficiency was assessed with a 50-item cloze test designed in-house, which included knowledge of vocabulary, grammar and reading comprehension. Proficiency was also evaluated by examining participants' accuracy on the comprehension questions included after each item in the main task, the Self-paced Reading Task (SPRT). Participants included in the analysis answered 76.6% of the 56 questions accurately on average (as shown in Table 1 below). After taking into account results from both tests, the data from 7 participants was discarded, because they lacked the proficiency level to perform the tasks included here. After the removal of those participants, the proficiency test and the accuracy data from the SPRT showed a significant correlation (r=.58). Table 1 shows the statistical summary for both proficiency measures.

Table 1

Descriptive statistics for proficiency measures

Tasks*	n	Mean	SD	Range
Proficiency Test	50	37.5	4.79	31-47
SPRT accuracy	56	42.9	8.27	27-54

* Tasks correlated with r = .58

4.2. Self-paced Reading Task

This was the critical measure in the study, designed to test the morphological sensitivity of the L2 learners when processing subject-main verb agreement relationships in Spanish. This task was implemented using E-prime (Schneider, Eschman & Zuccolotto, 2002), which allows for the recording of RT per word. There were three resulting measures from this test: 1) the reading time for each word presented in a sentence, 2) reading comprehension accuracy for each sentence read, and 3) reaction times for the comprehension questions that followed the stimuli.

4.2.1. Materials

There were 20 target items in this task with 36 filler sentences of an unrelated nature. The 20 target items were divided into two conditions: one in which the subject was immediately followed by the conjugated form of a main verb (Condition A), and another in which the verb was placed two regions/words away from the subject noun phrase (Condition B). The addition of distance between the noun phrase and the main verb in these sentences was meant to test whether the sensitivity to morphology was influenced by the immediacy between the elements involved in agreement (Foote, 2011; Keating, 2009) and whether distance played any role in the relationship between sensitivity and WM capacity. All sentences were made up of 9 regions or orthographic words. Each item within a condition presented a grammatical and an ungrammatical version, as shown in Tables 2 and 3. All sentences were followed by a comprehension question to keep participants focused on meaning.

Table 2

Stimuli for condition A

Regions	R1	R2	R3	R4	R5	R6	R7	R8	R9
Grammatical	La	perra	duerm***e***	en	el	garaje	cuando	hace	frío.
	The	*dog*	*sleeps*	*in*	*the*	*garage*	*when*	*it's*	*cold.*
Ungrammatical	*La	perra	duerm***o***	en	el	garaje	cuando	hace	frío.
	**The*	*dog*	*sleep*	*in*	*the*	*garage*	*when*	*it's*	*cold.*

Table 3

Stimuli for condition B

Regions	R1	R2	R3	R4	R5	R6	R7	R8	R9
Grammatical	El	cliente	del	restaurante	pid***e***	siempre	la	misma	ensalada.
	The	*customer*	*of the*	*restaurant*	*asks*	*always*	*for the*	*same*	*salad.*
Ungrammatical	*El	cliente	del	restaurante	pid***o***	siempre	la	misma	ensalada.
	**The*	*customer*	*of the*	*restaurant*	*ask*	*always*	*for the*	*same*	*salad.*

The regions of interest for Condition A were R2 (to discount any participant predictive strategies), R3 (the target region), R4, and R5 (to test for spillover effects). In Condition B, the regions of interest were R4, R5 (the target region), R6 and R7. If the participants in this study were morphologically sensitive, we would expect longer RTs for the target or spillover regions (as in Jiang, 2007) for the ungrammatical versions of stimuli in both conditions. If, on the contrary, learners were to replicate the performance of Chinese speakers in Jiang's study, there would be no significant RT differences between regions in either condition.

4.3. Automated Operation Span Task

The WM test was administered using a publicly-available version developed by Unsworth et al. (2005). In this computerized task, participants are presented with a simple math operation first, e.g., (1*2)+1= ?. After this, the following screen shows a number that may or not be the correct result for the math problem previously presented. The participant is then asked to make a True/False judgment, based on the number given as result on the screen. Once the participant's button press is recorded, the next screen shows a capital letter, which is the item to be remembered during the test. An important feature of this implementation is that during the test phase, the timing of the different presentations is controlled by the computer. This control is achieved by calculating the average time it takes each participant to solve the math equation and press a button to go to the true/false judgment on the following screen. The participant's performance is controlled by that average time, and when they take longer (the average plus 2.5 SDs) to solve an equation, the computer directly passes to the judgment screen, and it labels the trial as a speed error. The second processing measure involves the time it takes each participant to provide a judgment based on the number shown for the math problem. There are three sets of tokens of sizes 3-7 included in the test, for a total of 75 math problems and the corresponding letters to be remembered. After a set has been processed, the participant is shown a screen with a matrix of twelve letters and asked to choose only the letters that appeared during the set, in the order they were presented. The entire test is controlled using a PC mouse. Feedback is provided as to the number of letters recalled correctly after each set. The resulting measures for the task consist of the absolute number of letters recalled during the task (AOSTabs) and the total number of letters recalled in their correct position (AOSTtoc), both of which are measures of storage capacity. As mentioned above, the test also records the number of speed errors (AOSTsper) and the overall time taken to provide a judgment about the result of the equation (AOSTmath).

If WM plays a role in how sensitive L2 learners are to morphological mistakes during parsing, we would predict a significant positive correlation between RTs and WM scores on the target and spillover regions of the ungrammatical versions in conditions A and B of the SPRT stimuli. This means that participants who have a higher WM capacity would be able to detect, or show sensitivity to, the morphological mismatch in the target region, and thus take longer to read the target and spillover regions of these sentences.

4.4. Raven's Advanced Progressive Matrices Test

The Advanced version of the Raven's Advanced Progressive Matrices Test (RPMT) is a test of intellectual ability (especially fluid intelligence) which involves 36 items. As discussed above, it was included in order to increase the validity of the WM capacity instrument previously described. Each item presents an abstract image matrix with one section missing in the bottom right corner. The participant's task is to choose the missing piece from a set of 8 possible candidates. This is a paper-and-pencil test for which the participant is allotted 40 minutes to complete all of the 36 items. The resulting measure is the total number of items completed correctly. If the WM test described above is actually testing the storage and processing capacity of the central executive within Baddeley's model (Baddeley, 2003), we would predict a positive correlation between the AOST and the RPMT to independently verify the validity of the WM test used.

4.5. Procedure

Participants' proficiency in the L2 and their performance on the RPMT were assessed during regular class time as a group. The rest of the tests were administered individually during the second experimental session. The second session began with the SPRT for half of the participants tested, while the other half performed the AOST first. The SPRT allowed for a break toward the middle of the test and most participants made use of this option. Participants who successfully completed all of the tests were given extra credit in the Spanish class in which they were enrolled.

In order to process the RTs for all stimuli in the SPRT, a mean RT by participant was calculated and any value that exceeded the participant's mean by 2.5 standard deviations (SDs) was removed from the dataset and replaced with the upper limit value (mean + 2.5SDs) for that participant.

5. Results

Results are presented according to the order of the research questions outlined above. Hence, the SPRT reading profiles will be described first, since these measures will allow us to determine whether the L2 learners in the study were sensitive or not to the morphological markings (for person) presented in the sentences read using a moving-window display. This description will be followed by the scores obtained from the individual differences measures: the AOSPT and the RPMT. The details from these findings will prove useful when answering the question of whether higher WM capacity provides an advantage while processing morphology in an L2.

The RTs shown in Table 4 correspond to the regions of interest in the SPRT, with the corresponding SDs (provided below each raw RT) and RT differences between grammatical and ungrammatical versions (found under the SDs).

Paired samples t-tests showed that the learners' performance presented significant differences for Condition A (where the subject and main verb were immediately adjacent) on R3, the locus of the error, and R4, the first spillover region. A similar pattern obtained for the stimuli in Condition B: there was a significant difference in RTs between the grammatical and ungrammatical versions for R3, and a difference with a trend towards significance for the first spillover region, R4.

Table 4

Reading times by condition and sentence type for the SPRT

		Condition A				Condition B			
Sentence Type	n=18	R2	R3	R4	R5	R4	R5	R6	R7
Grammatical	Mean	738.14	714.67	666.53	604.54	788.08	748.2	715.07	702.65
	SD	(449.7)	(335.3)	(283.3)	(233.73)	(441.22)	(504.73)	(339)	(483)
Ungrammatical	Mean	703.15	1089.39	928.87	704	720.19	1012.35	849.51	688.29
	SD	(403.3)	(826)	(390.8)	(364.4)	(281.06)	(476.43)	(411.09)	(265)
	Diff.	-34.99	374.72*	262.34*	99.47	-67.89	264.15*	134.44#	-14.36

* significant at .05; # p = .109

Following Jiang's (2007) analysis of the data, Table 5 shows the paired samples t-tests for RTs according to region and condition. For Condition A, the comparisons involved the grammatical and ungrammatical versions of regions 2-5 in each of the target sentences. For Condition B, the t-tests involved comparisons between the grammatical and ungrammatical tokens of regions 4-7 for each of the sentences read.

Table 5

Paired-samples t-tests for RTs in the SPRT

	Comparisons	*t*	*df*	*Sig.*
Condition A	R2 GR. vs. UNGR	.755	17	.461
	R3 GR. vs. UNGR	2.45	17	.025
	R4 GR. vs. UNGR	4.153	17	.001
	R5 GR. vs. UNGR	1.124	17	.277
Condition B	R4 GR. vs. UNGR	.812	17	.428
	R5 GR. vs. UNGR	2.453	17	.025
	R6 GR. vs. UNGR	1.69	17	.109
	R7 GR. vs. UNGR	.132	17	.897

Figures 1 and 2 show the reading profiles for grammatical and ungrammatical stimuli for both conditions. The Y axis shows the raw RTs in milliseconds and the X axis is divided into the regions of interest.

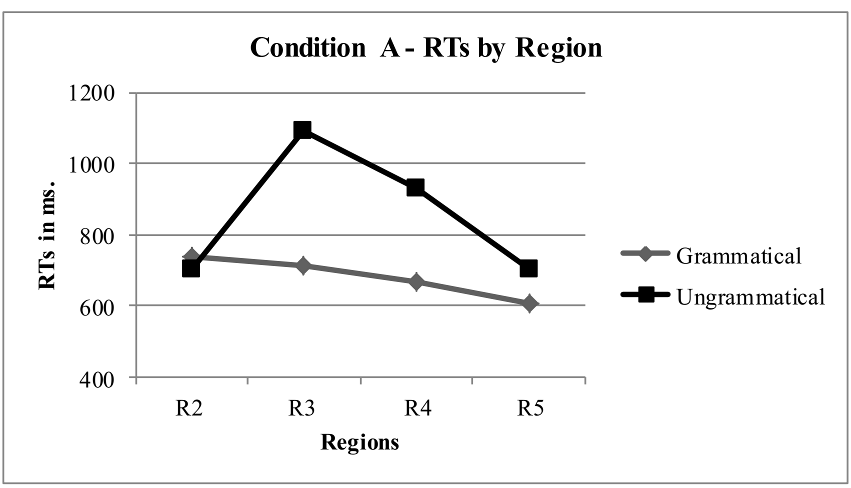

Figure 1. *RTs by region for condition A*

As can be seen in Figure 1, the ungrammatical stimuli in Condition A, which involved adjacency between the subject noun phrase and a main verb that disagreed in person with the subject, triggered a significant increase on RTs over the target region (R3), as well as on the first spillover region (R4). On the other hand, the grammatical sentences did not present any significant differences between regions.

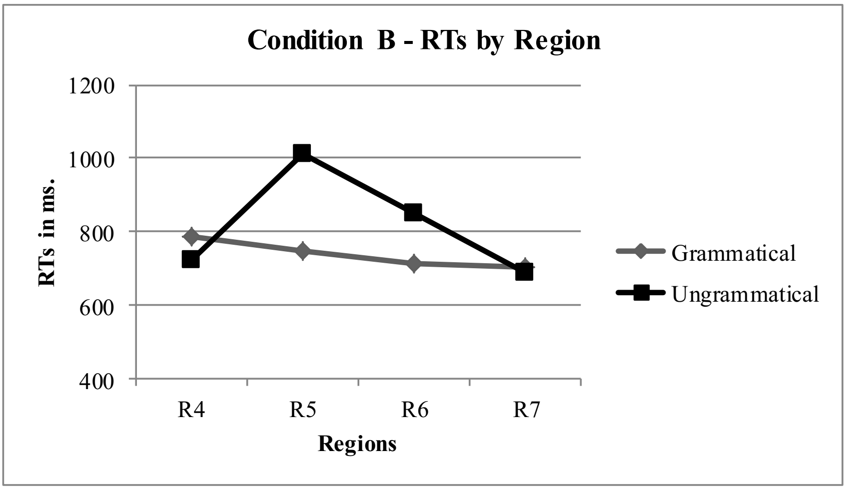

Figure 2. *RTs by region for condition B*

Figure 2 shows a similar pattern for sentences in Condition B, including sentences with a two-word/region distance between the subject noun phrase and the onset of the misconjugated main verb. In this case as well, the significant spike in RTs took place on the target region (R5), and the following region presented an increase that showed a probable tendency towards significance (with $p = .109$).

Table 6 below gives the mean and SDs for the individual differences tests used in the study.

Table 6

Descriptive statistics for individual differences tests

N = 18	Items	Mean	SD
AOSTabs	75	64.67	10.53
AOSTtoc	75	53.44	16.2
AOSTsper	75	.89	.83
AOSTmath	18	9132.33ms	4856.95ms
AOSTerr	75	4.72	2.08
RPMT	36	24.11	5.12

The AOST provides two measures of storage capacity (abs and toc), and it also allows for the measurement of three processing indicators: the number of speed errors (sper), the overall reaction time to the math operation (math), and the number of errors when solving these equations (err). Speed errors involve an item within a set in which the participant was not able to solve a simple math operation and press a button to continue to the result true/false judgment within the individual average time limit calculated during the practice session. The following Tables, 7 and 8, present the correlation matrices for the WM test and the fluid intelligence test, and their relationship to RTs in both stimuli conditions.

As expected with the AOST results, participants' absolute score (AOSTabs) was higher than the number of recalled letters in the right order of presentation (AOSTtoc). This feature of the results attests to the difficulty participants have when remembering items serially.

Table 7

Correlations for WM measures and RTs in the SPRT for condition A

				Grammatical Sentences				Ungrammatical Sentences			
	AOSTabs	AOSTtoc	RPMT	R2	R3	R4	R5	R2	R3	R4	R5
AOSTabs	1	.910**	-.130	.272	.251	.259	-.104	.238	.283	.171	.367
		.000	.606	.275	.316	.299	.682	.343	.255	.497	.134
	18	18	18	18	18	18	18	18	18	18	18
AOSTtoc	.910**	1	-.017	.138	.102	.206	-.294	.163	.260	.064	.308
	.000		.947	.585	.688	.412	.237	.517	.298	.801	.213
	18	18	18	18	18	18	18	18	18	18	18
AOSTerr	-0.441	-0.305	.042	-.294	-.291	-.217	-.367	-.136	-.387	-.287	-.233
	0.067	0.218	.869	.236	.242	.387	.135	.591	.112	.248	.352
	18	18	18	18	18	18	18	18	18	18	18
AOSTsperr	-.188	-.166	-.480*	-.107	-.228	-.016	.141	-.136	-.217	-.380	-.093
	.455	.512	.044	.674	.364	.951	.576	.592	.388	.120	.714
	18	18	18	18	18	18	18	18	18	18	18
AOSTmath	.178	.245	.211	.055	.116	.119	.183	.139	.567*	.275	.164
	.479	.327	.400	.830	.645	.637	.467	.582	.014	.270	.516
	18	18	18	18	18	18	18	18	18	18	18
RPMT	-.130	-.017	1	-.097	-.075	.081	-.058	-.080	.018	-.093	.096
	.606	.947		.702	.766	.750	.818	.752	.942	.713	.704
	18	18	18	18	18	18	18	18	18	18	18

** Correlations significant at the .01 level (2-tailed)

Table 7 shows the correlations between individual differences measures (the AOST and the RPMT) and the RTs for grammatical and ungrammatical sentences in Condition A of the SPRT, which involved a subject noun phrase and an adjacent main verb. As can be seen there, no relationship was found between the RTs for the critical regions and the storage measures in the WM test or the fluid intelligence instrument. What is more, there was no correlation between the WM scores and the results of the RPMT of abstract reasoning. The only significant relationship was the expected correlation between the AOST absolute score and the number of letters recalled in order (AOSTtoc).

Table 8

Correlations for WM measures and RTs in the SPRT for condition B

				Grammatical Sentences				Ungrammatical Sentences			
	AOSTabs	AOSTtoc	RPMT	R4	R5	R6	R7	R4	R5	R6	R7
AOSTabs	1	.910**	-.130	.061	.212	-.026	.024	.256	.236	.245	.217
		.000	.606	.809	.399	.919	.924	.306	.345	.327	.360
	18	18	18	18	18	18	18	18	18	18	18
AOSTtoc	.910**	1	-.017	-.034	.203	-.068	.065	.196	.236	.218	.374
	.000		.947	.894	.418	.788	.797	.435	.347	.385	.126
	18	18	18	18	18	18	18	18	18	18	18
AOSTerr	-0.441	-0.305	.042	-.317	-.421	-.020	-.007	.190	-.110	-.290	-.118
	0.067	0.218	.869	.199	.082	.936	.979	.449	.663	.243	.641
	18	18	18	18	18	18	18	18	18	18	18
AOSTsperr	-.188	-.166	-.480*	-.203	-.259	-.202	-.178	.212	-.095	-.116	.001
	.455	.512	.044	.419	.299	.422	.479	.398	.706	.645	.996
	18	18	18	18	18	18	18	18	18	18	18
AOSTmath	.178	.245	.211	.304	0.367	.432	.392	.021	.458	.418	.120
	.479	.327	.400	.220	.153	.920	.108	.934	.056	.084	.636
	18	18	18	18	18	18	18	18	18	18	18
RPMT	-.130	-.017	1	.154	.192	.354	.208	-.137	.014	.040	-.168
	.606	.947		.542	.445	.150	.408	.589	.955	.875	.506
	18	18	18	18	18	18	18	18	18	18	18

** Correlations significant at the .01 level (2-tailed)

A similar picture arises from the correlations obtained for the stimuli in Condition B, shown in Table 8. No relationships were found between RTs on the critical regions and the storage WM measures or the fluid intelligence test. However, the data show a negative correlation between the speed errors in the AOST and the RPMT (r=-.48, p=.044). There are also relationships to be noted between the AOSTmath (the time taken to solve equations) and the RTs for some of the regions in the ungrammatical stimuli. The AOSTmath measure correlates positively with the target region in Condition A (r = .567, p=.014), and two of the regions, R5 and R6, present a trend towards significance (r=.458, p=.056 and r=.418, p=.084 respectively) in Condition B.

6. Discussion

According to the research questions previously stated, the main objectives of this study were to investigate if morphological insensitivity was a universal characteristic of all L2 learners, or whether it varied according to the L1 of the participants and their individual cognitive capacities. An additional question focused on assessing whether the processing of person morphology on the verb also showed signs of insensitivity in L2 learners. The RT results from the SPRT showed that the English-speaking learners of Spanish tested in this experiment were indeed sensitive to morphological mismatches between a subject noun phrase and the corresponding main verb when reading sentences one word at a time in the target language. Contrary to what was found in Jiang's (2007) study of plural morpheme sensitivity among Chinese-speaking learners of English, the parsing profiles of these L2 learners showed increased RTs at the locus of the error in both conditions included here. Thus, sensitivity was present when the agreement relationship was adjacent and also when the matching words were separated by two intervening regions. This latter finding also runs contrary to previous results in the literature that had identified a difference in sensitivity based on distance between the members in the agreement relationship (Keating, 2009). Additionally, the sensitivity that the participants showed stretched over spillover regions in both conditions. These results regarding RTs show that the specific L1 that learners bring to the task may have a facilitative role (see also Foote, 2011 and Sagarra & Herschensohn, 2010) when parsing morphology accurately, since both English and Spanish mark the main verb for person (even though they do so with varying frequency). Conversely, the L1 grammars of participants in Jiang's original study did not share the morphological marking of the plural, given that Chinese uses this grammatical feature very sparingly (Sun, 2006). It may also be the case, as argued by DeKeyser (2000) that these grammatical markers across languages differ in their degree of salience, and, as such, person morphology in Spanish predicates is perceptually easier to identify than the plural –s marker in English. While this remains outside the scope of this paper, it represents an interesting avenue for crosslinguistic research in the future.

In terms of the influence of WM in morphological sensitivity, the findings here show no relationship whatsoever between the storage measures (abs and toc) in the AOST and the RTs in regions of interest, which is puzzling. When processing these kinds of agreement relationships, without automatic competence, learners would have to store the feature for person that corresponds to the subject and then reactivate that information when the main verb of the utterance is encountered. Thus, the storage capacity of WM would definitely be involved in at least the matching of person features for the subject and main verb. Nevertheless, neither measure of storage correlated with RTs in the SPRT. There were, however, a few interesting correlations between one of the measures of processing in the AOST, the overall reaction time in the math operation, and the target and spillover regions of ungrammatical stimuli in the SPRT. The positive correlations described above point to a relationship between the speed at solving simple math equations and the time spent processing a morphological mistake in the L2. The more time it took learners to solve a math operation, the longer they spent reading the target or spillover regions of ungrammatical stimuli, suggesting increased morphological sensitivity. This may indicate the application of a strategy while reading the ungrammatical regions of stimuli in the SPRT. However, it does not really shed light onto the reasons why no relationship was found between the other WM measures and the RTs for grammatical or ungrammatical regions. It may be the case though, that intermediate learners of Spanish, who have been instructed for more than 4 semesters in the target language, have automatized (DeKeyser, 2001) the verbal agreement marker for person in their L2. Consequently, greater WM capacity would not be of use if the structure or language component has already been integrated in the target language. In order to assess this hypothesis, it would have been useful to add a comparison group of native speakers of Spanish. Their performance in all tasks would have helped in determining whether the learners showed signs of automatization. Since morphology is thought to be automatized in native speakers, we would not expect any interactions between RTs and WM scores in the processing profiles of native speakers of Spanish, as WM capacity is not thought to exert any influence on knowledge that has already been integrated into the speaker's competence. Thus, if native speakers' and learners' scores showed no correlations, it would add support to the hypothesis that the intermediate learners in this study had automatized knowledge of person morphology when processing verbs in the target language.

Another unexpected finding as regards the WM results is the lack of a relationship between the AOST and the RPMT. Both tests have been shown to correlate in previous psychometric research (Wiley et al., 2011). However, neither the abs nor the toc values of the AOST correlated with the fluid intelligence measure. The author speculates this may have been due to the kind of implementation of the RPMT adopted for this study. We used a paper-and-pencil version of the RPMT which was timed—participants were given 40 minutes for the 36 items included in the test. The availability of more than a minute to complete each item may not have exerted sufficient cognitive demands, so that a relationship would emerge between the two instruments. In a different RPMT implementation, which involves only 18 items, participants were given 10 minutes to complete the task (Unsworth & Spillers, 2010). In that study, significant correlations between WM measures and the RPMT obtained.

In sum, the findings in the current study suggest that not all learners are morphologically insensitive. Moreover, when these findings are considered in the context of previous research, the capacity to process morphology effectively seems to vary according to both the morphological marker in question and the L1 of the learner. The issue of the influence of WM in this process remains unanswered in this study, mainly due to the fact that the learners assessed here seemed to have already automatized verbal morphology in their repertoire. It would be enlightening to use these same instruments and stimuli with learners of a lower proficiency level, since the lack of automatization may force such learners to recruit all their WM resources in order to process these sentences accurately. Another problem that seems pervasive in the field, and which makes it difficult to arrive at definitive claims about WM capacity, is the variety of tests used in the literature. It would be beneficial for the field to adopt a standard WM test, such as the one used here – the AOST – which is easily administered and imposes the same constraints on all users automatically.

In future research, the implementation of a shorter, timed, experimenter-controlled version of the RPMT may help in determining whether both measures are correlated in L2 learners. It would also be of value to extend Jiang's pioneering work to other language pairs with diverse morphological markers to test the construct of morphological insensitivity more exhaustively.

This new avenue of research involving the moment-to-moment assessment of what learners of different proficiency levels and L1s do with the morphology that receives so much attention in instructional settings might prove useful in bridging the gap between the research lab and the classroom. The information gained in these studies may help better prepare pedagogical materials to the processing needs of L2 learners, particularly in trying to facilitate and automatize morphological knowledge through, for example, materials for extensive reading outside of class.

References

Baddeley, Alan. (2003). Working Memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4* (10), 829-839.

Clahsen, Harald & Felser, Claudia. (2006). Grammatical processing in language learners. *Applied Psycholinguistics 27*, 3-42.

Conway, Andrew, Kane, Michael, Bunting, Michael., Hambrick, D.Zack, Wilhelm, Oliver, & Engle, Randall. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review, 12,* 769-786.

Daneman, Meredyth, & Carpenter, Patricia A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19*(4), 450-466.

DeKeyser, Robert. (1997). Beyond explicit rule learning: automatizing second language morphosyntax. *Studies in Second Language Acquisition 19*, 195-222.

DeKeyser, Robert. (2000). The robustness of critical period effects in second language acquisition. *Studies in Second Language Acquisition*, *22*, 499–533.

DeKeyser, Robert. (2001). Automaticity and automatization. In P. Robinson (ed.), *Cognition and second language instruction*. New York: Cambridge University Press.

Engle, Randall, Nations, John K., & Cantor, Judy. (1990). Is working memory capacity just another name for word knowledge? *Journal of Educational Psychology, 82,* 799-804.

Foote, Rebecca. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics, 32*, 187-220.

Gathercole, Susan E. (2007). Working memory and language. In G. Gaskell (Ed.), *Oxford Handbook of Psycholinguistics*. Oxford University Press.

Hopp, Holger. (2006). Syntactic features and reanalysis in near-native processing. Second Language Research, 22, 369-397.

Hopp, Holger. (2010). Ultimate attainment in L2 inflection: Performance similarities between non-native and native speakers. Lingua, 120, 901-931.

Jiang, Nan. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 603-634.

Jiang, Nan. (2007). Selective integration of linguistic knowledge in adult second language acquisition. *Language Learning, 57,* 1-33.

Juffs, Alan & Harrington, Michael. (2011). Aspects of working memory in L2 learning. *Language Teaching 44*(2), 137-166.

Keating, Gregory D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning, 59*(3), 503-535.

McDonald, Janet L. (2006). Beyond the critical period: processing-based explanations for poor grammaticality judgment performance by late second language learners. *Journal of Memory and Language 55*, 381-401.

Pearlmutter, Neal, Garnsey, Susan, & Bock, Kathryn. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41*, 427-456.

Rai, Manpreet K., Loschky, Lester C., Harris, Richard J., Peck, Nicole R., & Cook, Lindsay. (2011). Effects of stress and working memory capacity on foreign language readers' inferential processing during comprehension. *Language Learning 61(1),*187-218.

Raven, John, Raven, John C., & Court, John H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales, Section 1: General Overview*. San Antonio, TX: Harcourt Assessment.

Robinson, Peter. (1997). Generalizability and automaticity of second language learning under implicit, incidental, enhanced, and instructed conditions. Studies in Second Language Acquisition, 19, 223-248.

Sagarra, Nuria & Herschensohn, Julia. (2010). The role of proficiency and working memory in gender and number agreement processing in L1 and L2 Spanish. Lingua, 120(8), 2022-2039.

Schneider, Walter, Eschman, Amy, & Zuccolotto, Anthony (2002). *E-Prime reference guide*. Pittsburgh, PA: Psychology Software Tools Inc.

Sun, Chaofen F. (2006). *Chinese: A Linguistic Introduction*. Cambridge, England: Cambridge University Press.

Ullman, Michael (2001). The neural basis of lexicon and grammar in first and second language: The declarative/procedural model. *Bilingualism: Language and Cognition, 4*(2), 105-122.

Unsworth, Nash, Heitz, Richard, Schrock, Josef, & Engle, Randall. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498-505.

Unsworth, Nash. & Spillers, Gregory. (2010). Working memory capacity: Attention control, secondary memory, or both? A direct test of the dual-component model. *Journal of Memory and Language 62*, 392-406.

Wiley, Jennifer, Jarosz, Andrew, Cushen, Patrick & Colflesh, Gregory. (2011). New rule use drives the relation between working memory capacity and Raven's Advanced Progressive Matrices. *Journal of Experimental Psychology: Learning, Memory, and Cognition 37*(1), 256-263.

Witzel, Jeffrey, Witzel, Naoko, & Nicol, Janet (2012). Deeper than shallow: Evidence for structure-based parsing biases in L2 sentence processing. *Applied Psycholinguistics, 33,* 419-456.

Selected Proceedings of the 2011 Second Language Research Forum: Converging Theory and Practice

edited by Erik Voss,
Shu-Ju Diana Tai, and Zhi Li

Cascadilla Proceedings Project Somerville, MA 2013

Copyright information

ISBN 978-1-57473-458-4 library binding

Ordering information

Orders for the library binding edition are handled by Cascadilla Press.
To place an order, go to www.lingref.com or contact:

Cascadilla Press, P.O. Box 440355, Somerville, MA 02144, USA
phone: 1-617-776-2370, fax: 1-617-776-2271, sales@cascadilla.com

Web access and citation information