Elicited imitation as a measure of morphemic accuracy: Evidence from L2 Spanish

DONNA E. WEST*

State University of New York at Cortland

Abstract

This study measures whether number and type of morphemes in an elicited imitation string result in a greater number of modifications with L2 experience. Rationale is drawn from L2 working memory processing limitations at distinct levels of proficiency. 38 subjects (L2 Spanish university students) comprise three proficiency groups: beginning, undergraduate majors and graduate students. Number of morphemes was varied within each syllable count; and responses were either correct or modified (lexemically/inflectionally as deletions or substitutions). One or two way ANOVAs determined significance between mean proportions for each group. Findings indicate that increases in number of morphemes yielded significant differences, and that while the lowest proficiency group produced higher proportions of lexical deletions, the modifications made by more advanced groups were inflectional substitutions.

Keywords
elicited imitation, working memory, second language acquisition, Spanish, inflections, morphemic processing

1. Introduction

Particular to this inquiry is the measurement of implicit as opposed to explicit L2 knowledge across distinctive proficiency levels by means of oral elicited imitation (EI) testing. Chunking information, at the morphemic level and beyond, becomes more likely with L2 development given the additional wealth of input and output, together with changes in working memory (WM) and the
interlanguage (IL). According to Bley-Vroman and Chaudron (1994: 245), “...If the subjects’ grammar corresponds to the grammar used in producing the string, the imitation is more likely to be correct. Specific inaccuracies may point to specific differences between the subject’s grammar and the grammar of the target string.” Lust et al. (1987) and Lust et al. (1996) likewise support this claim. Increases in vocabulary coupled with morphosyntactic advances account for a greater likelihood of chunking beyond the phonological level and responses should reflect reconstructions indicating the level at which chunking is operationalized for different L2 proficiency groups. Moreover, “the limitations of working memory drive learners to construct higher order chunks out of lower order chunks” (R. Ellis 2008: 469). Ellis (2001) contends that individual differences appear to exist between L2 learners of similar developmental or experiential levels; nonetheless, he argues that such differences are small in comparison to differences which exist across developmental levels, especially if they are sufficiently different to warrant separation into skill groups. Mackey et al. (2002) are in accord—learners’ developmental stage is a factor in noticing phonological strings in WM. Mackey et al. (2002), however, cite some exceptions—that learners with less L2 experience, with greater than typical WM capacity, performed comparably or better than did some of their more advanced counterparts. This issue demands further investigation—establishing whether and to what degree proficiency level and increases in morphemic chunking in WM capacity are correlated such that an increase in one is associated with an increase in the other. Hameyer’s (1980: 21) claim underscores the need for further research into the types (inflectional, derivational) of morphemic, not syllabic, modifications across proficiency groups: “...Incorrect inflections and incorrect substitutions are the surest single indicators by which to distinguish proficiency groups.” The present study investigates whether L2 learners with less language experience are more likely to produce deletions of inflectional affixes, and whether more advanced L2 learners are more likely to produce substitution errors in view of their breadth of semantic and syntactic knowledge, and increased chunking in WM.

2. Rationale

Elicited imitation has a lengthy history. Its superiority to other methods is grounded in three rationales: (i) elicited imitation does not leave production of a particular grammatical structure or lexical item to chance; (ii) elicited imitation establishes a uniform base of linguistic input for all learner participants; and (iii) elicited imitation can measure implicit knowledge if it elicits reconstructions of stimulus sentences. “Elicited imitation may be advantageous over spontaneous speech sampling because it allows greater control over contextual variables than is possible in most naturalistic settings” (Corrigan 1982: 223).
Because the stimulus sentences are developed to elicit particular grammatical categories and relationships, and because they are identical and presented in an invariant sequence for all subjects, they permit robust comparisons across subject performance. Implementing time constraints has two purposes: to guard against planning used in explicit knowledge (Erlam 2009: 72), and suppressing rehearsal, ensuring that the response is not mere verbatim recall (Baddeley 2007: 51).

EI’s initial purpose was to determine the limits of short term memory (STM). STM refers to the ability to store and process a limited amount of information online for a limited amount of time, typically two or three seconds (Eysenck 2001: 160; Baddeley 2003: 675). Miller (1956) determined that the limit of STM is seven units, plus or minus two, although Simon (1974) determined that STM does not go beyond five chunks. Miller’s (1956) determination that STM consists of a greater number of chunks may represent the number of digits and not the number of linguistic chunks. Chunks within STM can be larger than a single digit/phoneme/syllable/morpheme, perhaps depending on the proficiency level. The issue of types and composition of “chunks” has, to date, been given little attention. Lewandowski and Murdock (1989) found that digits are more easily remembered in STM than are linguistic chunks. Erlam’s (2006) findings further suggest that linguistic units that are meaningful are more likely to be remembered in STM or WM; and Munnich et al. (1994: 237) indicate that modifications made to the stimulus sentence demonstrates meaningful processing. Since the 1970s, STM has been replaced by WM, extending its limitations in terms of functionality and capacity. Because WM is responsible for additional functionality and capacity when compared to STM, its means to hold additional chunks likewise surpasses that of STM (Skehan 1998: 50). Establishing the limits of STM and/or WM lays the groundwork for the use of EI as a measure of auditory memory units (cf. Skehan 1998: 78).

Levelt (1989) applies Baddeley’s WM model to issues of L2 learning and proposes that WM plays a primary role in L2 acquisition, especially at beginning and intermediate proficiency levels. Levelt proposes several specific L2 related functions for WM: accessing lexical or semantic chunks from LTM; altering LTM chunks to accommodate to the input or message; and assembling LTM chunks and WM chunks either simultaneously or in quick succession. Levelt’s model makes prominent three functions of language processing in WM: recall, modification, and assemblage. According to Lennon (2000),

---

1. According to Gathercole and Baddeley (1993: 25), “The most convenient and widely used index of this developmental increase is provided by auditory digit span, the maximum number of digits that someone can immediately remember and then repeat back in the same order” — typically measured by a prompt provided to the listener, and then having the listener within three seconds articulate what he or she remembers.
Oppenheim (2000) and Segalowitz (2000), lexical access is slower and more deliberate in L2 at lower and intermediate proficiency levels; and a focus on content words at beginning levels (VanPatten 2004: 8) may inhibit notice of inflections (Sagarra 2008: 134), resulting in fewer within-word and fewer multi-word chunks within WM. As a consequence, greater attentional resources must be expended on organizing the many chunks of inflectional (bound) morphemes than on integrating LTM chunks (based on meaning from lexical items) into the message or assembling/sequencing linguistic chunks once in WM. The latter function—sequencing morphosyntactic items in WM—is likely to be less utilized since attentional resources are expended on LTM retrieval and message coordination. This burden on attentional resources is responsible for an increased likelihood of L2 errors or slower serialized articulation (non-automatic speech) (Temple 1997: 87). N. Ellis (2008: 380–386; Ellis and Cadierno 2009: 11–12) attributes protracted acquisition of inflections to be a consequence of low salience and proactive inhibition, often from L1. The need to quantify attentional allocation of linguistic information in WM, especially for inflectional morphemes, militates in favor of EI tests as the most adequate measure in view of its increased control over input, and hence output, together with its assessment of implicit linguistic knowledge (Ellis 2005: 166, 2009: 61).

3. Literature review
EI is experiencing resurgence in the L2 literature because it is deemed to be a good measure of second language competence. In fact, EI as a measure of WM capacity may serve a more critical purpose in L2 than in L1 development especially for lower and intermediate proficiency learners since lexical retrieval is typically slower and more conscious than for native speakers (Lennon 2000; Oppenheim 2000; Segalowitz 2000) and since chunks are smaller, hence more numerous for early L2 learners, and may consist of lower order linguistic units, e.g. the phonological versus the grammatical or semantic.

Naiman (1974) was one of the first to use EI to measure L2 competence. First and second grade children learning L2 French produced identical structures in the EI experiment as in natural speech (tense markers and object nouns and pronouns) (Naiman 1974: 70). Naiman only vaguely alludes to the types of modifications or reconstructions that his subjects made to the EI prompts, and he does not focus on developmental issues. Moreover, since all of Naiman’s EI prompts were 15 syllables in length, the effect of syllable length or, more importantly, chunks (phonemic, morphemic, syntactic, and the like) on WM was not addressed. Although Naiman indicates that 15 syllables “overloaded

2. For recent EI literature reviews, see Vinther (2002) and Jessop et al. (2007).
Elicited imitation measuring L2 morphemic accuracy

STM” (Naiman 1974: 72), he does not establish its limits before and after immersion.

Hamayan et al. (1977) used EI to test slightly older children’s (8, 11) L2 competence with syntactic items, namely, compound and complex structures, and wh-questions. They report no between-group differences on any of the syntactic items, word order and the like, although mean scores were not provided. Perhaps the groups’ L2 developmental levels were not sufficiently distinct to reflect differences. Further, these investigators do not address the effect of increased morphemic length or prompt meaningfulness on EI performance.

To demonstrate grammatical competence, specifically that of anaphora, Munnich et al. (1994) constructed an EI test instrument consisting of grammatical and ungrammatical prompts in four categories—48 items with 12 items per category. Subjects include advanced adult ESL learners only; and the issue of delay between stimuli and repetitions is unaddressed. Although implicit WM constraints were considered, in that all prompts consist of 15 syllables, the structure of both the grammatical and ungrammatical prompts is obscure and ambiguous even for native speakers. Insurance of the meaningfulness of test items/prompts is paramount since without such subjects’ reconstructions may not be measured—but verbatim recall only.

In line with Bley-Vroman and Chaudron’s (1994: 256) observation, Munnich et al. (1994) unwittingly ignore the pivotal role of serial effect in the stimulus sentence. A perusal of their test instrument reveals that subject pronouns are always in the middle of the prompt (at least four syllables back from the concluding syllable) while all object pronouns appear as the last syllable. This design defect—not varying placement—confounds the factors that contribute to pronoun deletion. Hence, their finding that participants were more likely to correct or reconstruct ungrammatical subject pronouns is called into question—subject pronouns may have been corrected as a consequence of their invariant placement.

L2 reconstructions in Erlam (2006) were deliberately integrated as a design feature in her EI test—following oral administration of the stimulus sentences (grammatical/ungrammatical), a filled delay was implemented prior to the imitations, which the previously mentioned studies failed to include. Subjects were instructed to judge whether each stimulus sentence was true or false prior to the respective imitation. While the filled delay in imitations is necessary to preclude sub-vocal rehearsal and guard against parroting, thus increasing the likelihood of restructuring, the incorporation of truth-value judgments may have served as too great a distractor. Rather than encouraging a focus on meaning, hence encouraging reconstructive processes, truth-values are likely to have interrupted focus on meaning. This is in accord with Bialystok’s (2001: 139–151) claim that it is difficult to disentangle grammatical anomalies from semantic ones, especially for bilinguals. Such is attributed primarily to
increases in an already burdened and limited working memory. “If the sentence
[ungrammatical] also contains incorrect semantic information, then it becomes
more difficult for subjects . . . to ignore these errors and attend only to the well-
formedness criteria” (Bialystok 2001: 140). Determining truth value (founda-
tional upon semantic knowledge) is likely to be confounded with determin-
ing grammaticality, such that an ungrammatical sentence is often considered
“untrue” because it is ungrammatical. In fact, half of Bialystok’s ungrammatical
sentences which were semantically accurate were considered to be inaccurate
by nine-year-olds and beyond in their native language (Bialystok 1986: 506–
507). Furthermore, the binary nature of truth value provides a fifty-percent
likelihood of a correct guess, and does not provide substantive evidence of
meaningfulness.

A preferred approach might include employing only grammatical stimulus
sentences, and afterward asking “who, what, where, when, why, or how” short
questions about the stimulus string; since such requires a (more qualitative)
non-binary response, which more reliably demonstrates the meaning ascribed
to them. This approach, together with measuring types of morphemic process-
ing in WM vs. syllable processing, has the advantage of demonstrating which
types of grammatical categories are subject to substitution or deletion at par-
ticular L2 developmental levels, which may indicate the relative degree of
salience and/or processing difficulty in WM at each level. The working
hypotheses are as follows: (1) the beginning proficiency group will produce a
higher proportion of lexical deletions than the more advanced groups; (2) more
advanced proficiency groups will produce a higher proportion of lexical substi-
tutions than lexical deletions; (3) the more advanced groups will produce
greater proportions of inflectional substitutions than the beginning group; (4)
with an increase in morphemic count, there will be greater proportions of over-
all errors for all proficiency groups; and (5) increases in syllable count have no
effect on performance at any developmental level.

4. The study

4.1. Participants

Thirty-eight undergraduate and graduate L2 Spanish students voluntarily par-
ticipated in this study. The L1 for all students is English, although two students
are simultaneous Spanish/English bilinguals.3 Fifteen students belong to the

3. The simultaneous bilinguals were born in the United States, and were more dominant in
English, in that it was spoken both in their home and exclusively in school. Their performance
on the oral EI measure was sufficiently similar to that of the advanced non-simultaneous bilin-
guals in the pool. In fact, three of the non-simultaneous bilingual participants scored higher
accuracy levels.
beginning group (novice mid to novice high on the Simulated Oral Proficiency Interview [SOPI])⁴—they experienced a mean of 1.4 semesters of university foreign language coursework, and a mean of 2.3 courses in high school. This beginning group did not experience any other non-classroom exposure to Spanish. Seventeen students belong to the advanced undergraduate group (intermediate mid to intermediate high on the SOPI)—they experienced a mean of 3.7 university courses (two of which constitute Spanish major courses), and a mean of 2.8 courses in high school. Approximately one fourth of these advanced undergraduate students studied in the target culture for one semester. Six students belong to the advanced graduate group (advanced low to advanced high on the SOPI); they experienced a mean of 2.5 graduate level Spanish courses, 10 Spanish major undergraduate courses, 2.6 beginning/intermediate undergraduate target language courses and 4.0 Spanish courses at the high school level.

4.2. Materials

A set of 24 oral EI prompts in Spanish were developed from vocabulary and structures already covered prior to data collection from the sole text (Vistas) used at the novice and intermediate levels, and were evaluated by two native Spanish speakers to ensure comprehensibility.⁵ These 24 prompts were then piloted on 4 students from each level. All students heard the same recording (24 prompts) produced by a female native Spanish speaker, who articulated carefully so as not to delete or modify phones, as in the process of synalepha.⁶ The order of the prompts with respect to syllable length and morphemic complexity was randomly sequenced; but the order remained constant for all subjects. Complexity here refers primarily to the number of morphemes within an individual syllable (although the issue of zero morphology may confound this),

⁴ These proficiency levels were developed by the American Council on the Teaching of Foreign Language and represent one of the primary metrics of proficiency utilized to assess high school and university students in the United States, as well as for foreign language teacher certification.

⁵ Slight irregularities in structure or meaning from typical standard language patterns can be useful in EI stimulus sentences to elicit reconstructions (Erlam 2006, 2009); in other words, providing a stimulus sentence which slightly alters what is often heard is beneficial, since learners can compare such structure to representations in the interlanguage.

⁶ While this pattern may appear artificial for the more advanced students, its implementation for the beginning students was intended to facilitate comprehension and to guard against imperceptibility of syllables/morphemes, which could have resulted in still higher proportions of morphemic errors. This is especially important in Spanish since e.g. gender, person, and number often conclude the syllable and the word, and are thus more susceptible to deletion or elision.
and secondarily to the number of meanings associated with a single form. In Spanish, morpheme count can be increased without increasing the syllable count, e.g. \( \text{Lim-pi(a)} \ l(a) \ \text{ven-ta-n(a)} \ \text{an-ti-gu(a)} \) — the \(-\text{a}\) means infinitival vowel on the verb; and the \(-\text{a}\) on the determiner noun and adjective represents gender. The more complex morphological correlate additionally includes the \(-\text{s}\) familiar addressee verb marker, and the \(-\text{s}\) plural markings on the remaining NP forms: \( \text{Lim-pi(a)(s)} \ l(a)(s) \ \text{ven-ta-n(a)(s)} \ \text{an-ti-gu(a)(s)} \), each consisting of nine syllables, but eight and twelve morphemes respectively. At minimum two prompts separated prompts with identical lexemes according to Valian and Aubry’s (2005) design, e.g. \( \text{Entiende el título extranjero} \) (‘He/she understands the foreign title’) was presented at least two prompts before or after \( \text{Entiendes los títulos extranjeros} \) (‘You fam. understand the foreign titles’). Intervening interference is a factor which contributes to the decay of articulatory representation in WM (Valian and Aubry 2005: 635). Hence, this procedure clears verbatim or meaningful traces which would remain in WM, were lexically similar prompts to be contiguously (or nearly contiguously) presented.

All of the prompts consist of four words: a determiner, noun, adjective, and present tense verb with differing syntactic sequences altering the placement of the verb. The adjective always followed the noun, and the determiner always preceded it, which is the typical sequence for descriptive adjective placement in Spanish. Subject pronouns were excluded, which conforms to typical Spanish patterns, the affix on the verb is the only person marker in those instances. Ten of the stimulus sentences included subject NPs which precede the verb, two included object NPs which precede the verb; and twelve object NPs which follow the verb. Half of the verbs were third-person singular in form, which are not represented morphemically, while an additional morpheme consisting of one additional phoneme was added to the other half, e.g. \(-\text{s} / -\text{n}\), second-person addressee and third-person plural, respectively. Person, number, and gender inflections were selected to test morphemic processing in WM, as tense

8. Because word order in Spanish is naturally more flexible than in English, the sequence variations did not result in awkward or contrived clausal structures. The variation in NP and VP sequence was deliberately orchestrated as to not confound memory of the lexical category with primacy or recency effects in WM — issues which are not directly addressed here. If subject NP is always first, and/or object NP is always last, memory of either constituent could demonstrate memory preferences for certain lexical categories as opposed to memory for initial or final structures.
9. Rationale for using second-person singular and third-person plural inflections (and not first-person) resides in the fact that the latter: (a) encodes a greater array of meanings in the form, e.g. \(-\text{o}\) indicating first-person, singular number, and present tense; and (b) requires syllable count increases, e.g. \( \text{to-ma-mos} \) (‘we drink’).
and aspect often increase syllable count, and are overly advanced for novice learners. Person, number, and gender represent frequent forms in the input to novice learners alike.

Number of syllables and number of morphemes is varied without varying the word count. Syllable length ranges from 7–11 with 8 prompts at each syllable count. This author chose 11 syllables since the ceiling in WM for L2 intermediates is 15 syllables, the ceiling for novices, although undetermined, is lower than that of L2 intermediates. Although Erlam extended syllable length to 18, her subjects’ L2 level was at minimum intermediate.

All stimulus sentences are declarative present. Determiners and nouns constitute subjects within ten of the prompts; the subject is marked as a verb inflection in the remaining fourteen. Placement of the verb within the string was varied (12 verbs at the conclusion and 12 at the outset).

The number of morphemes in the prompts ranges from six to twelve; the differential in each pair is typically four morphemes (see Appendix A). Half the prompts contain either a second person -s or the -n plural person marker on the verb, 5 and 7 tokens respectively, while the remaining 12 prompts are singular third-person forms. All verbs with singular subject NPs accordingly were third-person singular present tense; all verbs in prompts for which subject NPs were plural were present tense third-person plural, or present tense second-person singular to increase the number of morphemes without increasing syllable length.

Half of the nouns are masculine, and half feminine, and half of each of those end in -e which does not distinguish gender. Likewise, the number of gender morphemes was controlled across prompts—3 per prompt; while number morphemes amount to 4 per prompt, including determiners, nouns, adjectives, and verb affixes.

4.3. Procedure

Participants individually heard the same instructions in English and were instructed to repeat exactly what they heard and as much as they heard. They were likewise instructed that they would be asked to answer a short question pertaining to the meaning of each prompt before repeating the prompt. The questions ranged from two to three short words; the subjects were told that their answers needed to consist of only one to two words. Example stimuli, questions, and responses include:

10. As the ceiling for morpheme count in WM is undetermined, this range was selected to conform to numbers of units/chunks in WM to be well below the ceiling for native speakers (seventeen units), in light of the novice learners.
La niña buena llega (‘The good girl arrives’)
¿Cuándo llega? (‘When does she arrive?’)
Ahora. (‘Now.’)

Pides los pescados salados (‘You fam. order the salty fish’)
¿Para quién? (‘For whom?’)
Para mí. (‘For me.’)

If a subject failed to repeat more than two entire prompts, the data were excluded from analysis.

A ‘second pass’ of the lexical items within the stimulus sentences (plural after singular/singular after plural, with at least two different prompts intervening) rests upon Valian and Aubry’s (2005) double presentation design.

4.4. Coding

A coding scheme accounting for correctness/incorrectness of morphemes and syllables was employed (see Appendix B). The coding scheme likewise accounted for whether a morpheme/syllable was deleted or substituted and whether the substitution was lexical or inflectional. Data were coded within two months of collection. A second individual was trained extensively in the use of the coding scheme. In fact, the primary investigator and the independent coder simultaneously coded 40% of the data, and agreed on 95% of their coding decisions.

5. Results

The procedures described in this study were successful in eliciting imitations from L2 learners of differing proficiency levels. A two-way ANOVA was conducted to test for differences in error types (inflectional substitutions, lexical deletions, and lexical substitutions) across the three proficiency groups. Inflectional deletions were not included as a category, since it is infrequent that suffixes in the form of inflections are deleted, leaving the root bare. The main effect for proficiency was not significant $F(2,35) = .24, p = .80, \eta^2 = .01$. However, the main effect for error type was both significant and large $F(2,70) = 25.83, p = .001, \eta^2 = .43$. As reflected in Figure 1, all three proficiency groups produced a greater percentage of lexical deletions, as compared to the percentages of inflectional errors and lexical substitutions (see Table 1 for the standard deviations). However, the significant and moderately large interaction between the proficiency and error type ($F(4,70) = .2.94, p = .026, \eta^2 = .14$) suggests that the pattern of errors differed across proficiency group.
A series of planned contrasts was conducted to test the hypotheses predicting specific patterns of errors for the three groups, using a method described by Rosenthal et al. (2000).

The contrast associated with the hypothesis that the BU group would produce a greater proportion of lexical deletions than would the AU and AG groups was found to be both significant and large, $t(35) = 3.14, p < .004, r^2 = .22$. A similar contrast reflecting a greater proportion of inflectional errors for the AU and AG groups as compared with the BU group was also found to be significant and moderately large, $t(35) = 2.45, p = .019, r^2 = .15$. Interestingly, no apparent difference between the three proficiency groups was present for the proportion of lexical substitutions, as reflected in Figure 1.

A two-way ANOVA was conducted to test the effects for level of proficiency and morphemic complexity (7, 9, or 11 morphemes). As expected, the main effect for proficiency was significant and large, $F(2,35) = 7.62, p = .002, \eta^2 = .30$. A series of planned contrasts for linear and non-linear effects revealed that the linear effect reflected in Figure 2 was significant and large ($t(37) = 3.53, p = .001, r^2 = .25$) with the number of errors declining with proficiency. However, neither the main effect for morpheme or the interaction between morpheme and proficiency were significant, $F(2,70) = 1.253, p = .297, \eta^2 = .07$.

A similar approach was adopted to examine the effects for level of proficiency and number of syllable errors. No significant main effects were found for number of syllable or proficiency, $F(2,70) = .10, p = .903, \eta^2 = .003$ and $F(2,35) = .53, p = .593, \eta^2 = .03$, nor was a significant interaction found, $F(4,70) = .88, p = .482, \eta^2 = .05$.

Table 1. Mean proportions of errors and standard deviations

<table>
<thead>
<tr>
<th></th>
<th>BU (n = 16)</th>
<th>AU (n = 16)</th>
<th>AG (n = 6)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
</tr>
<tr>
<td>Inflection</td>
<td>0.11</td>
<td>0.06</td>
<td>0.27</td>
</tr>
<tr>
<td>substitutions</td>
<td>0.68</td>
<td>0.14</td>
<td>0.52</td>
</tr>
<tr>
<td>Lexical deletions</td>
<td>0.23</td>
<td>0.15</td>
<td>0.22</td>
</tr>
<tr>
<td>Morpheme (7)</td>
<td>0.47</td>
<td>0.12</td>
<td>0.31</td>
</tr>
<tr>
<td>Morpheme (9)</td>
<td>0.54</td>
<td>0.34</td>
<td>0.42</td>
</tr>
<tr>
<td>Morpheme (11)</td>
<td>0.38</td>
<td>0.19</td>
<td>0.18</td>
</tr>
<tr>
<td>Syllable (7)</td>
<td>0.21</td>
<td>0.12</td>
<td>0.67</td>
</tr>
<tr>
<td>Syllable (9)</td>
<td>0.45</td>
<td>0.20</td>
<td>0.22</td>
</tr>
<tr>
<td>Syllable (11)</td>
<td>0.59</td>
<td>0.12</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Note: BU—Beginning undergraduate students, AU—Advanced undergraduate students, AG—Advanced graduate students. Means for inflection substitutions, lexical deletions, and lexical substitutions represent mean of proportion of total errors. Means for morpheme and syllable reflect mean proportion of errors for each morpheme or syllable length (7, 9, or 11).
Finally, a series of exploratory analyses were conducted to examine the relationship between proportion of morpheme-related errors (7, 9, and 11 collapsed) and those discussed earlier (inflectional substitutions, lexical deletions, and lexical substitutions). Preliminary correlations between these variables revealed a significant positive correlation between morpheme errors and lexical deletions ($r = .51, p = .001$), and a significant negative correlation between morpheme errors and inflectional substitutions ($r = -.50, p = .001$). The correlation between morpheme errors and lexical substitutions was neither significant nor large ($r = -.17, p = .31$). To examine the extent to which morpheme errors may account for the effects for the other three error types, the initial two-way ANOVA (error type by proficiency) was conducted anew with the percentage of morpheme errors entered as a covariate. The results revealed a significant interaction between error type and the proportion of morpheme errors,
Elicited imitation measuring L2 morphemic accuracy

Once morpheme errors were controlled for, however, the main effect for error type and the interaction between error type and proficiency were non-significant and small, $F(2,70) = 2.311, p = .107, \eta^2 = .06$ and $F(4,70) = 1.253, p = .297, \eta^2 = .07$ respectively.

6. Discussion

Hypothesis 1 is supported, since the beginning undergraduates produced higher proportions of lexical deletions as compared to the two more advanced groups. It is interesting that as overall errors increased in morpheme production, so too did the proportion of lexical deletions, which may indicate that stimulus sentences with greater morphemic count provide greater opportunity for the production of lexical errors, namely deletions (see the discussion regarding Hypothesis 4 for elaboration). Although lexical deletions decreased with developmental level, lexical substitutions did not, disconfirming Hypothesis 2—that lexical substitutions would increase with developmental level. In fact, no significant difference between any of the three groups was found with respect to lexical substitution. This finding indicates that for these stimulus sentences, substituting lexemes was not more likely at higher proficiency levels, despite the presumably increased vocabulary and knowledge of words which share semantic attributes. The assumption was that the latter skill would advantage the more advanced L2 groups in that rather than deleting a word which they were unable to remember, advanced L2 learners would replace it with another to maintain the morphemic and syllabic counts, as well as the meaning, and to ‘fill the slot’ with a term having similar features. Instead, what appears to be operating is that rather than deleting lexemes, the more advanced L2 learners remember them more accurately, especially if they constitute roots as opposed to suffixes.

Hypothesis 3 is supported, since suffixes, in the form of inflectional morphemes on verbs and nouns, were treated differently for the more advanced proficiency groups, but not for the beginner group. Since inflectional errors (in the form of substitutions) comprise a greater proportion of the more advanced L2 learners’ errors when compared to the beginner group, it appears to be the case that the beginning L2 learners were less likely to process each lexeme at the morphemic level. The beginner group may not have noticed or may not have been as aware of the grammatical categories such as person, number, et cetera, as compared to the more advanced groups. As a consequence of a failure to morphemically decompose the individual lexemes from the stimulus string in WM, the beginner group was likely to make inflectional substitutions, while because the more advanced groups availed themselves of morphemic decompositions in WM, they did not chunk each lexeme in a formulaic manner. Hence, inflectional substitutions constituted a more likely error type.
Hypothesis 4, that all three proficiency groups would produce greater proportions of error types combined with increases in morphemic count, is supported. And as predicted, Hypothesis 5 is likewise supported, demonstrating that syllable count had no effect on performance at even the most basic proficiency level. What appears to have contributed most to error production for all three groups is lexical deletions. It is suspected that recency effects played substantially into this performance pattern. Were the data analyzed to consider primacy and recency effects in WM, which would be experienced by all proficiency levels, the lexemes in the middle would be less likely to be remembered, and hence be deleted. And, the pattern of lexical or inflectional substitutions may substantially take place in the initial or final lexemes of the stimulus sentences.

The advanced groups’ familiarity with the language permits them to make meaningful substitutions/reconstructions. In fact, subjects of the two more advanced groups substituted more sensically than did subjects of the beginner group, e.g. substituting a noun having similar syllable structure, gender, and number for one which they could not remember and/or a substitution which shares some semantic features, such that the substituted lexeme is sensical in the context. In other words, the original lexeme and the substitution can function equally well syntagmatically, e.g. entiende pájaro extranjero (‘He/she understands the foreign bird’) for En-tien-des-los-tí-tu-los-ex-tran-je-ros (‘You fam. understand the foreign titles’) or pide el pescado delgado (‘He/she orders the thin fish’) for pi-de-el-pe-sca-do-sa-la-do (‘He/she orders the salty fish’). It appears that the higher the linguistic level of control in WM, the more likely the learner is to substitute meaningfully where gaps in memory exist. When more advanced learners operate on the morphemic level and beyond, they are able to recover or hold meanings in WM while retrieving reasonable substitutions from LTM into WM for those morphemes temporarily suppressed or not readily retrievable. This operation on the part of more advanced L2 learners materializes simultaneously with maintenance of the phonological and syllabic structure likewise within WM. In contrast, beginner L2 responses contained fewer lexemes than did those of the advanced L2 learners, indicating that the beginners experience difficulties holding phonological and lexical information in WM while accessing a suitable substitution from LTM. Essentially, these less proficient learners appear to experience challenges switching from the phonological loop to LTM to access the substitution and integrating both in WM. Such requires coordination of an automatic nature. If the coordination between the phonological loop and information from LTM (accomplished by the episodic buffer) were not automatic, additional resources on an already limited WM system would need to be called upon (Baddeley 2000: 421, Baddeley 2007: 146–151).

The significant difference between the beginner and undergraduate advanced groups validates the primary claim of this study that it is the number of mor-
Elicited imitation measuring L2 morphemic accuracy

phemes within the string which operates to distinguish proficiency groups, especially obvious between the lowest and more moderately proficient groups.11 Although proportion of inflectional errors/reconstructions increased with proficiency level, overall morphemic errors decreased with proficiency level. Nonetheless, once morphemic errors were controlled for, an interaction between error type and proficiency was not found, i.e. the beginner and the more advanced L2 learners increased production of lexical deletions and of inflectional substitutions respectively, these no longer reached significance levels. These findings suggest that once morpheme count overburdened WM for any of the proficiency groups, these error types each reached significant levels. When morpheme counts within the stimulus sentences were moderate to low, perhaps at 9 or below, the error types remained constant for all three groups. Furthermore, the influence of morpheme count on errors is not only quantitative, but qualitative—the counts of both are positively correlated; and only when WM resources are expended for each group do these error types (lexical deletions vs. inflectional substitutions) materialize.

When WM became overloaded, the beginner group appears to have deleted not merely the inflection, but the entire lexeme, more often than substituting the inflections. In sum, the significantly lower proportion of inflectional errors on the part of subjects of the lowest proficiency level appears to be a consequence of two factors: (1) a greater dependence on verbatim recall rather than on reconstructive processes, and (2) the lack of WM resources at early developmental stages to notice or call up from LTM morphemic units within a string, especially suffixes. This latter rationale is in line with Ellis’s (2006) claim that bound morphemes, especially suffixes, are less noticeable as a consequence of their placement and their abbreviated phonetic duration, e.g. -s referring to number within the NP or person within the VP consists of but one phonemic unit and often appears in the middle of a string.

The more advanced groups are operating on the morphemic level and beyond, while the level of linguistic control of the lowest proficiency L2 learners appears to be primarily phonemic, and often a consequence of verbatim recall. Moreover, the less developed interlanguage of the lowest proficiency L2 learners may account for their reliance on the phonetic distribution of the stimulus sentence rather than on reconstructing its morphological/semantic/syntactic components. It is this factor (reconstruction at higher levels of linguistic control) which appears to be responsible for the modifications made to inflections when WM is overloaded.

What requires further investigation is how L2 maturational factors may influence the type of ‘error’ produced such that beginning L2 learners produce

11. For a more detailed explanation and rationale for the validity of this argument and its particular application to L2 Spanish, see West (2010a, 2010b).
greater proportions of within constituent alterations (e.g. gender, number), while more advanced L2 learners produce a higher proportion of across constituent alterations (e.g. person). Such analysis could uncover more dramatic performance differences between lower and higher proficiency L2 learners consequent to morphemic factors than was evidenced in this study.

7. Conclusion

Findings from this study suggest the need to use short and/or simple syntactic/morphemic structures within the stimulus sentences, such that performance of the beginner and intermediate proficiency groups can be adequately compared to that of the more advanced proficiency groups. Although some motivational/attention-based fallout can surface for the more advanced groups with this approach, without it performance comparisons between beginning L2 learners and more advanced L2 learners are unlikely to be measurable. Increasing the number of morphemes while holding the number of syllables in the string constant is a practice in need of further implementation.

The findings from this study further suggest that the number of morphemes in the stimulus string is a primary factor in distinguishing the proficiency groups, not the number of syllables. More specifically, findings from this study indicate that an increase in number of morphemes had a different effect (on the lowest proficiency group) increasing lexical deletions, while increasing inflectional substitutions for the more advanced L2 learners.

Appendix A: Stimulus sentences

<table>
<thead>
<tr>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>El viaje extensor encanta.</td>
<td>The lengthy trip is enjoyed.</td>
</tr>
<tr>
<td>Comen los postres ricos.</td>
<td>They/you (pl.) eat the delicious desserts.</td>
</tr>
<tr>
<td>El vinagre morado alimento.</td>
<td>The wine vinegar is nourishing/nourishes.</td>
</tr>
<tr>
<td>Los vinos tintos caen.</td>
<td>The red wines fall.</td>
</tr>
<tr>
<td>Toma la leche fría.</td>
<td>He/she drinks the cold milk.</td>
</tr>
<tr>
<td>Limpías las ventanas antiguas.</td>
<td>You (fam.) clean the old windows.</td>
</tr>
<tr>
<td>Pide el pescado salado.</td>
<td>He/she orders the salty fish.</td>
</tr>
<tr>
<td>Entiendes los títulos extranjeros.</td>
<td>You (fam.) understand the foreign titles.</td>
</tr>
<tr>
<td>La noche oscura regresa.</td>
<td>The dark night returns.</td>
</tr>
</tbody>
</table>
Appendix B: Coding scheme

Part of speech
Noun = 1
Verb = 2
Adjective = 3
Adverb = 4
Preposition = 5
Article = 6
Other = 7

Error

<table>
<thead>
<tr>
<th>Main type</th>
<th>Code</th>
<th>Sub 1</th>
<th>Code</th>
<th>Sub 2</th>
<th>Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deletion</td>
<td>0</td>
<td>Partial</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Deletion</td>
<td>0</td>
<td>Full</td>
<td>2</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Deletion</td>
<td>0</td>
<td>Morpheme</td>
<td>3</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Correct</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Lexical substitution</td>
<td>2</td>
<td>Unintelligible</td>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>
### References


Elicited imitation measuring L2 morphemic accuracy


