The storage and composition of inflected forms in adult-learned second language: A study of the influence of length of residence, age of arrival, sex, and other factors

LAURA BABCOCK, JOHN C. STOWE, CHRISTOPHER J. MALOOF, CLAUDIA BROVETTO and MICHAEL T. ULLMAN

DOI: 10.1017/S1366728912000053, Published online:

Link to this article: http://journals.cambridge.org/abstract_S1366728912000053

How to cite this article:

Request Permissions : Click here

Downloaded from http://journals.cambridge.org/BIL, IP address: 141.161.13.52 on 10 Sep 2012
The storage and composition of inflected forms in adult-learned second language: A study of the influence of length of residence, age of arrival, sex, and other factors

Laura Babcock
Brain and Language Laboratory, Department of Neuroscience, Georgetown University, Washington, DC, USA & Cognitive Neuroscience Sector, SISSA (Scuola Internazionale Superiore di Studi Avanzati), Trieste, Italy

John C. Stowe
Brain and Language Laboratory, Department of Neuroscience, Georgetown University, Washington, DC, USA

Christopher J. Maloof
Brain and Language Laboratory, Department of Neuroscience, Georgetown University, Washington, DC, USA

Claudia Brovetto
Brain and Language Laboratory, Department of Neuroscience, Georgetown University, Washington, DC, USA & ANEP (Administración Nacional de Educación Pública), Montevideo, Uruguay

Michael T. Ullman
Brain and Language Laboratory, Department of Neuroscience, Georgetown University, Washington, DC, USA

(Received: July 15, 2011; final revision received: November 7, 2011; accepted: January 19, 2012; first published online 30 April 2012)

It remains unclear whether adult-learned second language (L2) depends on similar or different neurocognitive mechanisms as those involved in first language (L1). We examined whether English past tense forms are computed similarly or differently by L1 and L2 English speakers, and what factors might affect this: regularity (regular vs. irregular verbs), length of L2 exposure (length of residence), age of L2 acquisition (age of arrival), L2 learners’ native language (Chinese vs. Spanish), and sex (male vs. female). Past tense frequency effects were used to examine the type of computation (composition vs. storage/retrieval). The results suggest that irregular past tenses are always stored. Regular past tenses, however, are either composed or stored, as a function of various factors: both sexes store regulars in L2, but only females in L1; greater lengths of residence lead to less dependence on storage, but only in females; higher adult ages of arrival lead to more reliance on storage. The findings suggest that inflected forms can rely on either the same or different mechanisms in L2 as they do in L1, and that this varies as a function of multiple interacting factors.

Keywords: late-learned second language (L2), length of residence (LoR), age of arrival (AoA), declarative memory, procedural memory

Introduction

Adult-learned second language (L2) has been the focus of much recent research, which has begun to elucidate important aspects of the neural and computational (neurocognitive) bases of L2 (for reviews, see Abutalebi, 2008; Clahsen, Felser, Neubauer, Sato & Silva, 2010; Doughty & Long, 2005; Gass & Mackey, 2012; Gor, 2010; Green, 2003; Hernandez & Li, 2007; Indefrey, 2006; Kotz, 2009; Kroll & de Groot, 2005; Morgan-Short & Ullman, 2012; Paradis, 2009; Schmidt & Roberts, 2009; Steinhauer, White & Drury, 2009; Ullman, 2005). Nevertheless, substantial gaps in our knowledge remain. Of interest here is the fact that it is still unclear whether L2 depends on the same or different neurocognitive bases as first language (L1), and whether such (dis)similarity varies as a function of factors such as language domain (e.g., lexicon vs. grammar), length of exposure to the L2, age of L2 acquisition (i.e., within adulthood), similarity

Address for correspondence:
Michael Ullman, Brain and Language Lab, Department of Neuroscience, Georgetown University, Box 571464, Washington, DC 20057-1464, USA
michael@georgetown.edu
between the L2 and the learner’s native language, and intrinsic subject-level differences such as sex (male vs. female). In this study we attempt to shed light on aspects of these issues.

A number of theoretical perspectives addressing these issues have been put forward. These perspectives can be divided into three broad groups. One suggests that the neurocognitive substrates underlying L2 are basically the same as those underlying L1. This group of theories can in turn be split into those that assume that lexical and grammatical functions depend on distinct neurocognitive systems (Abutalebi, 2008; Indefrey, 2006), and those that eschew this distinction (Ellis, 2005). Since, according to these theories, L1 and L2 always rely on the same neurocognitive substrates, the factors listed above should not affect the extent to which this holds true (although of course such factors could affect the difficulty of learning the L2). A second group of theories posits that the neurocognitive underpinnings of L2 are fundamentally distinct from those subserving L1 (Bley-Vroman, 1989). Again, factors such as those mentioned above should not have any effect on this relation.

The third group of theories takes an intermediate position between the other two. These theories hypothesize that L2 learners initially depend largely on different substrates than do L1 speakers, but, with increasing exposure or proficiency, gradually come to rely on L1 neurocognition (Clahsen & Felser, 2006b; Paradis, 2009; Ullman, 2001). The nature of this increased reliance differs among the particular theories. Paradis (1994, 2004, 2009) argues that this reliance increases across both lexical and grammatical functions. In contrast, both Clahsen and Ullman and their colleagues posit that lexical functions are always qualitatively similar in L1 and L2, and that only aspects of grammar change over time regarding the neurocognitive mechanisms they depend on (Clahsen & Felser, 2006a, b; Ullman, 2001, 2005, 2012). Clahsen argues that morphological and syntactic processing in L2 depend more on lexical, pragmatic, and world knowledge than in L1 (Clahsen & Felser, 2006b; Silva & Clahsen, 2008). However, “word-level [grammatical] processing and morphosyntactic feature matching between adjacent or locally related words might be more easily mastered as grammatical proficiency increases and can eventually become native-like” (Clahsen & Felser, 2006b, p. 565). According to Ullman, complex forms that depend largely on rule-governed compositional mechanisms in L1 rely instead on different mechanisms at lower levels of L2 exposure: in particular, declarative memory, which also underlies lexical memory (Ullman, 2001, 2005, 2006, 2012). This dependence on declarative memory may be instantiated in several ways, including the storage of complex forms as chunks (e.g., English forms like walked or the cat) and the explicit learning of rules. (Here we do not take a position as to how such chunks are represented, e.g., structured or unstructured, analyzed or unanalyzed.) However, Ullman posits that with increasing exposure, rule-governed aspects of grammar may increasingly be learned in the procedural memory system. This system is thought to subserve much of grammar in L1, and is posited to underlie the real-time composition of complex forms (e.g., walk + -ed, the + cat). This group of theories is also open in principle to the possibility that various other factors, in addition to exposure or proficiency, might affect the attainment of L1-like neurocognition (Clahsen & Felser, 2006a, b; Ullman, 2005, 2012). Clahsen has argued against such a role for the factor of L1–L2 similarity (Clahsen & Felser, 2006a, b); Ullman has not taken a position on this factor. Neither Clahsen nor Ullman have taken positions regarding the role of adult age of acquisition. Ullman has suggested that sex may play a role in at least L1, and possibly L2, with women relying more on declarative memory for rule-governed complex forms (e.g., by chunking), relative to men, due to a female advantage at declarative memory and possible disadvantage at procedural memory (Prado & Ullman, 2009; Ullman, 2004, 2005; Ullman, Miranda & Travers, 2008).

Different lines of research have elucidated different aspects of the neurocognitive relation between L1 and L2, and the factors that affect it. Here we focus on the computational nature of lexical and grammatical functions, in particular on the distinction between storage and composition. This distinction has been intensively studied in L1 in the contrast between regular and irregular inflectional morphology, especially in English past tense morphology (McCelland & Patterson, 2002; Pinker & Ullman, 2002). Whereas regular past tenses are rule-governed (add -ed to the stem), and may be composed, irregular past tenses are idiosyncratic (e.g., bring—brought, sing—sang, fling—flung) and so must depend (at least in part) on some sort of stored lexical representation. Examining regular and irregular forms thus allows one to contrast rule-governed grammatical (and potentially compositional) and idiosyncratic lexical aspects of language, while holding other factors constant (e.g., item-level factors such as the phonological length of inflected forms, or task-related factors).

Different patterns of the composition and storage of regular and irregular inflected forms are expected by different L2 theories. Those that posit that L2 and L1 depend on the same neurocognitive mechanisms (Abutalebi, 2008; Ellis, 2005; Indefrey, 2006) would expect the same pattern of storage and composition in L1 and L2. Those that assume different substrates for L1 and L2 (Bley-Vroman, 1989) might expect different patterns. In contrast, theories that hypothesize a shift from non-L1 to L1 substrates (Clahsen & Felser, 2006b; Paradis, 2009; Ullman, 2001) expect that the pattern of storage...
and computation in L2 will eventually come to resemble L1, but might not initially. Both Clahsen and Ullman have made specific predictions along these lines: whereas irregular inflected forms should be stored in both L1 and L2, regulars should tend to depend on lexical/semantic memory (e.g., chunked) at low exposure or proficiency levels, while at higher levels they have an increasingly likelihood of being composed by the grammatical system (Clahsen & Felser, 2006a, b; Ullman, 2001, 2005). Moreover, certain factors, such as sex (Ullman, 2005), may modulate this shift from storage to composition (e.g., if females are better than males at declarative memory and possibly worse at procedural memory, L2 females may initially depend more on chunking, and L2 males more on composition).

Previous evidence

Previous evidence regarding the storage vs. composition of regular and irregular past tense and other inflected forms has come largely from two methodological approaches: the examination of frequency effects and of priming effects.

Frequency effect studies are premised on the following logic. A set of forms that is stored in memory will show frequency effects – that is, forms with a higher frequency of occurrence in the language are processed more quickly or accurately than lower frequency forms. In contrast, composed forms (e.g., walk + -ed) will not show such “full-form” (or “surface”) frequency effects (e.g., past tense frequency effects) because such forms are not accessed as full forms in memory. Thus, frequency effects can be taken as a diagnostic for storage, with their absence suggesting composition, if one holds potentially confounding factors constant (for an alternate interpretation of an absence of frequency effects, i.e., by single-mechanism models, see Discussion and conclusions). Across multiple studies of L1 speakers, and over different languages, irregular inflected forms quite consistently show frequency effects, whereas regular inflected forms do not (Alegre & Gordon, 1999; Pinker & Ullman, 2002; Prado & Ullman, 2009). Rather, various item- and subject-level factors seem to influence whether or not regulars show frequency effects. For example, frequency effects on regularly inflected forms have been found to be more reliable in females than males (Prado & Ullman, 2009). Overall, these findings have been taken to suggest that irregulars always depend on stored lexical representations, while regulars are generally either composed (by a grammatical system) or chunked (in lexical memory), as a function of various factors (Alegre & Gordon, 1999; Clahsen & Felser, 2006b; Pinker & Ullman, 2002; Prado & Ullman, 2009; Ullman, 2004, 2008).

A number of studies have examined frequency effects in inflectional morphology in L2, across different languages. Three of these studies reported an absence of frequency effects for regularly inflected forms in L2 speakers. Lalleman, van Santen and van Heuven (1997) gave a Dutch past tense production task (dependent measure: response times) to L1 and “advanced” (quotation marks in this section indicate proficiency-related terms used in the cited studies) L2 users of Dutch (varied native language backgrounds; length of residence, and ages of first exposure and arrival not reported). They found frequency effects on irregular but not regular past tenses in both the L1 and L2 speakers. In an English past tense production task, Beck (1997) examined frequency effects (dependent variable: response times) in L1 and L2 speakers, over several experiments. The L2 speakers (varied native languages; mean length of residence (LoR): around three years in all experiments; age of first exposure to English and age of arrival in the U.S. not specified) showed the same pattern as L1 speakers on regulars, namely either an absence of frequency effects or anti-frequency effects, across the experiments. Irregulars (which were only tested in one experiment in L1 and L2) showed the expected frequency effects in L1, but not in L2 (an anomalous result). Birdsong and Flege (2001) examined frequency effects in regular and irregular English past tense and plural forms in a multiple choice task (dependent measure: accuracy) in L2 English speakers, who were considered to be “at L2 asymptote” (half were Spanish native speakers and half were Korean native speakers; LoR: 10–16 years; age of arrival: 6–20 years). They reported an interaction between frequency (high vs. low) and regularity (regular vs. irregular): whereas irregular past tense and plural forms showed frequency effects, regulars did not. The Spanish and Korean learners of English did not differ with respect to their frequency effect pattern. L1 English speakers were not examined.

In contrast, three studies have reported frequency effects for regularly inflected forms in L2 speakers. In a study of Spanish present tense production (dependent measure: response times), Bowden, Gelfand, Sanz and Ullman (2010) found differing patterns of frequency effects for L1 and L2 speakers (English learners of Spanish; mean of about nine months in immersion contexts; age of first immersion exposure above 17 years), whose mean proficiency was “Advanced Mid” (SOPI; Center for Applied Linguistics, Washington, DC). The L1 Spanish speakers showed frequency effects for both regular (e.g., vender–vendo “to sell – I sell”) and irregular (e.g., perder–perdo “to lose – I lose”) present tense forms of class I/II/III verbs (whose infinitives end in -er/-ir), and for irregulars (e.g., pensar–pienso “to think – I think”) of class 1 (-ar verbs), but not for class I (the default class) regulars (e.g., pescar–pesco “to fish – I fish”). The L2 speakers, however, exhibited frequency effects on regular and irregular present tense forms in all cases, including
for regulars from class I. Neubauer and Clahsen (2009) probed German past participles in a lexical decision task (primary dependent measure: response times) given to L1 speakers and highly “competent” L2 German learners (Polish native speakers; mean LoR: 3.1 years; mean age of first exposure: 14.5 years). The native speakers exhibited frequency effects on irregular but not regular past participles, while frequency effects were found on both types of verbs in the L2 users. In another lexical decision study (primary dependent measure: response times), Silva (2009) examined regular past tense forms in L1 and “advanced” L2 English speakers (Chinese and German native speakers; mean LoR: 14.8 months; mean age of first exposure: 13.7 years). Irregulars were not examined. While both the L1 and L2 speakers showed frequency effects for the regulars, the effects were larger for L2 users. The Chinese and German native speakers did not differ in their frequency effects. An examination of sex differences in this study revealed no differences between males and females within any of the L1 or L2 groups.

In sum, previous studies of frequency effects have shown the following. A few studies have reported a lack of frequency effects for regularly inflected forms in L2 learners (Beck, 1997; Birdsong & Flege, 2001; Lallemant et al., 1997), suggesting the (de)composition of regulars in these subjects. These L2 speakers appeared to have substantial exposure and/or high proficiency in two studies (Birdsong & Flege, 2001; Lallemant et al., 1997) though apparently not in a third (Beck, 1997). In contrast, in three other studies, the pattern of frequency effects suggested the storage of regularly inflected forms in L2 learners but not (or less so Silva, 2009) in L1 speakers (Bowden et al., 2010; Neubauer & Clahsen, 2009). In these studies, the L2 learners had relatively low LoRs or immersion exposure, though their proficiency levels seemed to vary somewhat across the studies (possibly due in part to the heterogeneity of proficiency measures used). Unlike regulars, most studies (all but Beck, 1997) that examined irregularly inflected forms found frequency effects in L2, as well as in L1 where examined (Birdsong & Flege, 2001; Bowden et al., 2010; Lallemant et al., 1997; Neubauer & Clahsen, 2009). Thus, previous frequency effect studies suggest that idiosyncratic inflected forms depend on stored representations in both L1 and L2, whereas regular inflected forms vary in their computation. They can be either stored or composed, as a function of various factors, including sex (in L1) as well as the amount of exposure for L2 learners: regulars appear to be (de)composed at least to some extent in L2 learners with higher exposure (or proficiency), whereas in learners with lower L2 exposure they tend to be stored.

In priming studies that attempt to distinguish between the storage and composition of inflected forms, participants are typically presented with inflected forms followed by related forms of the same word (e.g., stem or infinitival forms). The demonstration of full priming (e.g., regularly inflected forms facilitate the target as much as when the target is primed by itself, i.e., as much as identity priming) is generally interpreted as suggesting decomposition, whereas its absence is taken to suggest a lack of decomposition, implying storage. We are aware of two studies that have employed this paradigm in L2 learners. (Here we focus on studies that test for full priming, and thus do not discuss other types of priming paradigms: Basnight-Brown, Chen, Hua, Kostic & Feldman, 2007; Feldman, Kostic, Basnight-Brown, Durdevic & Pastizzo, 2010). Silva and Clahsen (2008) examined English past tense priming in L1 and “proficient”/“competent” L2 speakers (Chinese, German, and Japanese native speakers; mean LoR: 16.1 months; mean age of first exposure: 13.3 years). They found that whereas the L1 speakers showed full priming for regulars (e.g., equivalent priming for walked–walk and walk–walk), the L2 speakers did not (regulars did not prime their stems any more than did unrelated primes). This L2 pattern did not differ between the Chinese and German native speakers (who were not compared to the Japanese speakers). Irregulars were not examined. The same pattern was found in a study of German inflectional morphology (Neubauer & Clahsen, 2009) that examined L1 and “competent” L2 speakers of German (Polish native speakers; mean LoR: 3.1 years; mean age of first exposure: 14.4). For the L1 speakers, regular past participles fully primed (regular) first person singular present tense forms (i.e., the past participles primed these forms as much as they primed themselves). In contrast, the L2 speakers did not show such priming for the regulars (which did not prime the present tense forms any more than did unrelated primes). Unlike the regulars, irregular past participles showed partial priming (the inflected form primes the target less than the target primes itself, but more than an unrelated form), in both the L1 and L2 speakers. In sum, these two L2 priming studies suggest that the L1 speakers decomposed regularly inflected forms, whereas the L2 speakers – who had low LoRs in both studies – did not. In contrast, the results from the one study that examined irregularly inflected forms suggest that these were not decomposed in either L1 or L2.

Together, previous studies of frequency and priming effects suggest the following. Whereas irregularly inflected forms always depend on stored representations, both in L1 and L2, regularly inflected forms can be either composed or stored. In L1, they are often but not always composed, with storage/composition apparently varying as a function of sex and other factors. In L2, the evidence suggests that regulars might be composed at higher LoRs and proficiency, but tend to be stored at lower LoRs (with the effects of proficiency not clear). Overall, this pattern is most consistent with the predictions of Clahsen
and Ullman. However, various methodological concerns in previous studies, as well as a number of empirical gaps, weaken our ability to draw a clear picture of the nature of the storage and composition of inflected forms in L2 as compared to L1, in particular with respect to the factors of interest laid out above. These methodological concerns include: a lack of comparison of L2 effects with L1 (Birdsong & Flege, 2001); heterogeneous native languages across L2 subjects (Beck, 1997; Lalleman et al., 1997); and the absence of clear information about length of residence or other length of exposure measures (Lalleman et al., 1997), age of first exposure (Beck, 1997; Birdsong & Flege, 2001; Lalleman et al., 1997), or age of arrival where applicable (Beck, 1997; Lalleman et al., 1997; Neubauer & Clahsen, 2009; Silva, 2009; Silva & Clahsen, 2008). Moreover, no previous study has reported all three of these L2 measures (length of residence/exposure, age of first exposure, age of arrival where applicable).

In addition, there are several empirical gaps relevant to the present study. None of the studies presented above directly examined the effect of LoR or proficiency on the storage/composition of inflected forms, and thus did not directly test the predictions of Clahsen and Ullman. Similarly, the effect of age of arrival or age of first exposure on storage/composition has not been examined. In fact, in several studies, the age of arrival or of apparent initial substantial exposure was below age 17 (Birdsong & Flege, 2001; Neubauer & Clahsen, 2009; Silva, 2009; Silva & Clahsen, 2008), making it difficult to draw conclusions about adult L2 acquisition, the issue of interest here. Finally, sex differences were investigated in only one study (Silva, 2009), and just three contrasted storage/composition between L2 learners of different native languages (Birdsong & Flege, 2001; Silva, 2009; Silva & Clahsen, 2008).

Thus, the evidence to date seems to suggest that regularly inflected forms tend to be initially stored (chunked) but can eventually come to be composed by L2 learners. However, this shift in reliance from storage to composition as a function of exposure or proficiency has not been tested directly, holding other factors constant, let alone in direct comparisons with irregular forms and with L1 speakers. Moreover, the modulation of storage/computation by age of acquisition (e.g., as measured by age of first exposure or age of arrival), sex (male vs. female), and native language of the L2 learner remains unclear.

**The present study**

This study is designed to elucidate aspects of these issues. The study examines regular and irregular past tense frequency effects in an English past tense production task, in which response times of correctly produced forms constitute the dependent measure. We focused on English past tense because it is likely the best studied and best understood regular/irregular inflectional contrast, with which, moreover, we have substantial experience both in L1 and in the investigation of sex differences (Pinker & Ullman, 2002; Prado & Ullman, 2009; Ullman, 1999, 2004, 2008; Walenski, Mostofsky & Ullman, 2007; Walenski, Weickert, Maloof & Ullman, 2010). Thus, it provides a foundation of understanding on which we can build in the study of L2. The task was given to L1 speakers and L2 learners of English. The L2 participants had a mean LoR of 7.3 years, with a range of 3–23 years. Their mean age of arrival (AoA) was 27.0 years, with a range of 17–41 years. None of the L2 subjects had any substantial exposure to English before age 17 (see Participants in Material and methods for a definition of substantial exposure). About half of the L2 subjects were native speakers of Chinese, and half native speakers of Spanish. About half of the L1 group and of each L2 subgroup (Chinese, Spanish) was male, and half female.

The study contrasted past tense frequency effects in L2 and L1 for regulars and irregulars, and examined whether native language (Chinese vs. Spanish), sex (male vs. female), LoR, and (adult) AoA impacted the pattern of frequency effects. We examined frequency effects rather than priming effects because we have extensive familiarity with the former (Bowden et al., 2010; Prado & Ullman, 2009; Ullman, 1999; van der Lely & Ullman, 2001) but not the latter, and because frequency effects directly test for storage, allowing us to probe for the presence or absence of chunking for regulars. Spanish and Chinese native speakers were selected because Spanish has complex suffix-based inflectional morphology, while Chinese has little inflectional morphology (Chu, 1982; Henne, Rongen & Hansen, 1977), allowing us to contrast the influence of languages with inflectional systems of different complexity on regular past tense forms in L2 English. We focused on LoR rather than proficiency because LoR is more easily comparable across studies (whereas almost every study uses different proficiency measures). Moreover, previous studies of L2 storage/composition suggest that LoR might have a greater influence than proficiency on the attainment of composition for regular inflected forms (see above). Finally, LoR and proficiency are, nevertheless, generally highly collinear, and thus hard to tease apart. Therefore we did not acquire proficiency measures, and do not examine their role here (see Discussion and conclusions for potential limitations of this approach).

Our selection of subjects with no substantial exposure to English before age 17 allowed us to focus on AoA, which provided a measure of the onset of adult L2 acquisition for which subsequent exposure occurred in a relatively naturalistic context. The examination
of subjects with a wide range of LoRs and AoAs provided sufficient variability in these factors to examine their impacts on frequency effects. The analyses used mixed-effects regression modeling, which provides, among other benefits, the opportunity to include both item- and subject-level covariates in the same model, thus allowing one to simultaneously account for a wide range of potentially confounding factors. Indeed, the analyses controlled for 14 item-level and 12 subject-level covariates, first by testing whether they were statistically warranted for inclusion in the model, and then by actually including them if they were.

Our predictions were based on Ullman’s theory, the declarative/procedural (DP) model of L2 and L1, together with previous findings where this model is agnostic (Ullman, 2001, 2005, 2006, 2012). Frequency effects were expected for irregular past tense forms in L1 and L2 subjects, with no qualitative impact on these effects by any of the factors being examined, that is, native language, sex, LoR, or AoA. Frequency effects for regular past tense forms were expected to vary according to at least some of these factors. L2 speakers were expected to show stronger frequency effects on regulars than L1 speakers, all other factors being held constant. The DP model has not made specific predictions regarding the role of native language in L2 learners. Although the studies cited above suggest that native language does not play a role in the storage/composition of regular (or irregular) inflected forms in L2 (Birdsong & Flege, 2001; Silva, 2009; Silva & Clahsen, 2008), other studies of morphology are consistent with such a role (Bassnight-Brown et al., 2007; Feldman et al., 2010; Portin, Lehtonen, Harrer, Wande, Niemi & Laine, 2008), making strong predictions for this factor difficult. On the basis of the DP model and previous findings, females were predicted to show stronger frequency effects on regulars than males. This was expected in L1, but not necessarily in L2, since both male and female L2 learners might chunk regulars. Regular past tense frequency effects were predicted to decrease with increasing LoR, reflecting the gradual proceduralization of the rule-governed regulars. We had no specific predictions with respect to the role of adult AoA on the pattern of regular past tense frequency effects, since changes during adulthood in the relative functionality of the declarative and procedural memory systems (and thus changes in storage vs. composition) remain unclear (Ullman, 2005).

Material and methods

Participants

We tested 135 healthy adults. All participants were right-handed (Oldfield, 1971), had no known personal or family history of neurological, psychiatric, or learning disorders, and had at least 12 years of formal education (starting with 1st grade). All participants gave informed consent, and were paid for their participation. Participants completed a questionnaire, just prior to testing, which asked for a range of detailed information, including age, education, handedness, medical history, language background, and (for the second language learners) motivational and affective factors.

Of the 135 participants, 72 were native speakers of American English (L1 speakers), and 63 were non-native speakers of English (L2 learners). Somewhat similar analyses to those presented here were carried out on the participants reported by Prado and Ullman (2009), who constitute the L1 (control) group here. Of the 63 L2 learners of English, 30 were native speakers of Chinese and 33 were native speakers of Spanish. The L2 participants were all late learners of English, selected according to the following criteria. First, all of the L2 participants moved to the U.S. or another English-speaking country (e.g., England, or English-speaking Canada) for the first time at the age of 17 years or older. Second, none of them had had “substantial exposure” to English before their arrival in an English-speaking country. “Substantial exposure” is defined here as one or more of the following: (a) attending a bilingual school, with English as one of the languages, for more than one year; (b) living in an English-speaking country for more than six months; (c) having had more than five years of formal English instruction with classes every work/school day, for more than three hours per day; or (d) having had contact with native English speakers on a regular basis for more than one year (e.g., hosting an English-speaking student at home). Additionally, none of the participants’ “primary caretakers” (defined in the questionnaire as “parents, nannies, etc.”) were English-speaking (all were native Chinese or Spanish speakers), and none had prior English immersion experience in a non-native English-speaking society (e.g., Hong Kong, India). Finally, all participants must have lived in the U.S. for the three consecutive years prior to testing in this study.

Nine participants were excluded from analysis. First, a later review of the participants’ questionnaires revealed that seven L2 participants did not meet the criteria for inclusion. Six had not lived in the U.S. for the three consecutive years prior to testing, and one had lived and regularly spoken English in Hong Kong for 12 years prior to coming to the U.S. Finally, two L1 participants had past tense production response times greater than 2 standard deviations from the mean for L1, calculated over all 58 analyzed real verbs (see below) (see Prado & Ullman, 2009). Thus, analyses were performed on 70 L1 and 56 L2 participants (see Table 2 below).
Verbs

Each participant was tested on 112 verbs: 32 irregulars (e.g., hold–held); 32 “consistent” regulars (e.g., fail–failed), whose stems are not phonologically similar to the stems of irregulars; 16 “inconsistent” regulars (e.g., squeeze–squeezed), whose stems are phonologically similar to the stems of irregulars, and whose past tenses have an increased likelihood of storage (Prado & Ullman, 2009; Ullman, 1999); and 32 novel verbs (e.g., plag). Only the irregular and consistent regular verbs are analyzed and discussed here. The test items were selected to exclude no-change verbs (e.g., hit–hit) and doublet verbs, i.e., verbs that can take two past tense forms (e.g., dive–dived/dove).

To quantify frequency (not only past tense surface frequency, the primary independent variable of interest, but also stem surface frequency and lemma frequency), two frequency counts were used in combination: (i) the Francis and Kucera (1982) count, derived from a one-million-word corpus composed of a variety of sources; and (ii) a frequency count extracted by a stochastic part-of-speech analyzer from a 44-million-word corpus of unedited Associated Press news wires from 1988 (Church, 1988; Ullman, 1999). All analyses were carried out on the natural logarithm (ln) of the sum of the two frequency counts, first augmented by 1 to avoid the undefined ln(0) (Prado & Ullman, 2009; Ullman, 1999).

The 32 regular and 32 irregular verbs were matched pairwise on stem (unmarked form) frequency (transformed frequency values, pairwise t-test: t(31) = 0.08, p = .94), lemma frequency (sum of the frequencies of all verb forms, except past tense but including past participle: t(31) = 0.35, p = .73), and past tense surface frequency (t(31) = 0.79, p = .44). As in other studies using the same past tense production task (e.g., Prado & Ullman, 2009), three problematic verbs were excluded from analyses: sink is a doublet verb, with two possible irregular past tenses (sank and sunk); the past tense of bind is also a verb stem (bound); and owe was the only single-phoneme verb, and is phonologically similar to many more irregular verbs than any other verb selected as a consistent regular. The three regular or irregular verbs matched pairwise to these problematic verbs (sign, drown, and stride, respectively) were also removed from all analyses (Prado & Ullman, 2009). The remaining 29 pairs of regulars and irregulars were analyzed. Like the full set of 32 pairs, these 29 regulars and 29 irregulars did not differ on stem frequency (transformed frequency values: regulars: range = 1.10–8.59, $M = 5.68, SE = 0.36$; irregulars: range = 1.79–9.03, $M = 5.88, SE = 0.40$; paired $t(28) = 0.78, p = .44$), lemma frequency (regulars: range = 4.37–10.04, $M = 7.28, SE = 0.31$; irregulars: range = 3.85–9.85, $M = 7.42, SE = 0.32$; paired $t(28) = 0.65, p = .52$), or past tense frequency (regulars: range = 1.61–9.43, $M = 5.81, SE = 0.38$; irregulars: range = 1.79–10.19, $M = 5.81, SE = 0.40$; $t(28) = 0.01, p > .99$).

Procedure

The procedure followed in this study was the same as in other studies using this task; for further details, see Prado and Ullman (2009). Each verb stem was presented visually alone and in the context of a sentence, with a second sentence containing a blank to elicit the past tense form (e.g., fail. Every day I fail an exam. Just like every day, yesterday I ____ an exam). The verb stem and the two sentences were displayed at the same time on the computer screen, one below the other. The sentences for the different verbs were identical apart from the verb itself and the post-verbal material (e.g., fail and an exam). All post-verbal complements or adjuncts were composed of two words, neither of which was inflected or of low frequency. Items were pseudo-randomized, and all participants were presented with the items in the same order, to facilitate experimenter coding (see below for the inclusion of item order as a covariate to account for order effects). Participants were instructed to produce the missing form as quickly and accurately as possibly. To familiarize participants with the task, they were first given 10 practice items: three regulars, three irregulars, and four novel verbs. The entire test session was audio recorded on minidisk for later transcription. Accuracy was determined by the experimenter and an independent transcriber, both of whom were native English speakers. In rare cases of disagreement, a third native English speaker made the final determination. Response time (RT) data were recorded via a microphone connected to a computerized timer.

Analysis

Past tense frequency effects were examined only on response time data, as a high rate of correct responses among the L1 speakers precluded sufficient variability for analyses on accuracy (Prado & Ullman, 2009). Response time analyses were performed only on correct first responses (81.8% of all first responses to the 58 regular and irregular verbs). During testing, the experimenter noted items for which the computerized timer was not triggered by the onset of the participant’s response; these response times were excluded from analysis (1.7% of correct responses). RTs faster than 500 ms were discarded as being likely due to computer error (0.2% of correct responses). Extreme outliers for each participant—that is, responses whose RTs were more than 3.5 standard deviations from the given participant’s mean for all real verbs—were also excluded from the analysis (0.7% of correct responses). Significance of all effects was assessed
using $\alpha = 0.05$. All $p$-values are reported as two-tailed. In all analyses, degrees of freedom were computed using the Satterthwaite approximation.

The response time data were analyzed using mixed-effects regression models, with random effects of subject. This statistical method allows each individual response time from each participant and item to be entered into one model, without averaging response times, which results in a substantial loss of information. Moreover, it allows various item- and subject-level covariates that may influence the pattern of results to be included in the same model. Mixed-effects models account for participant variability by including the baseline performance (model intercept) of each participant as a random variable. For more complete discussions of this statistical method and for similar analyses see especially Baayen (2004), as well as Prado and Ullman (2009).

A mixed-effects model was constructed with natural logarithm-transformed response time as the dependent variable, past tense surface frequency and verb-type (regular vs. irregular) as item-level variables, and native language (English, Spanish, Chinese) and sex (male vs. female) as subject-level variables. Interactions were specified between native language, verb-type, sex, and frequency, generating separate frequency coefficients for English, Chinese, and Spanish speakers, separately for males and females, on regular and irregular items.

A number of potentially confounding item- and subject-level variables were examined for possible inclusion as covariates. Each of these variables was included as a covariate only if it met certain specific conditions that suggested it might confound the results. Thus, only a subset of potential covariates was actually included in analyses, reducing the risk of overfitting the data.

We examined 14 item-level variables (see Table 1) (Prado & Ullman, 2009) and 12 subject-level variables. For both native and second language speakers, the subject-level variables included age (in years) at testing and years of formal education (starting with 1st grade); see Table 2. For L2 speakers only, the following additional 10 variables were examined. First of all, following other studies of L2 (Birdsong, 1992; Birdsong & Molis, 2001; Flege, Yeni-Komshian & Liu, 1999; Johnson & Newport, 1989), we considered: age of arrival in the United States or another English-speaking country, age of first exposure to English (defined as the age at which the speaker first started learning English, with either formal instruction or immersion), and length (years) of residence in the United States or other English-speaking countries (see Table 3). Following Stevens (1999), length of residence was natural-log transformed in all analyses (e.g., because the first year of exposure is likely to have a greater impact than the 10th or 20th year), though note that it is not displayed as log-transformed in Table 3. We also examined three measures of formal English language education: years of formal English instruction prior to arrival in an English-speaking country; years of such instruction after arrival; and total years, including both before and after arrival. While many previous studies have considered only total years of instruction (Birdsong, 1992; Birdsong & Molis, 2001; Johnson & Newport, 1989), it is possible that formal instruction before and after arrival contribute differently to second language acquisition. Note that although all three of these measures of formal English language education were considered as potential covariates, only total years of instruction is shown in Table 3. Finally, we also considered four variables related to attitude and motivation regarding English (Table 3), as these factors have been hypothesized to also influence second language learning and performance (Birdsong & Molis, 2001; Flege et al., 1999; Johnson & Newport, 1989). Subjects were asked to rate themselves on a scale of one to five, with five being the highest, in their identification with United States culture, their self-consciousness at learning or using English in the United States, their motivation to learn English (“Is it important for you to be able to speak English well?”), and their desire to stay in the United States.

The following steps were taken to determine whether each variable would or would not be included as a covariate in analyses. First, it was assessed whether each potential covariate was associated with response time. Association was tested with separate correlations (Pearson’s $r$) between each potential covariate and either the mean subject scores (for the 12 subject-level covariates) or the mean item scores (for the 14 item-level covariates), with mean RTs computed as the mean natural logarithm transformed response times. A covariate was considered to be associated if it predicted performance ($p \leq .05$) over all subjects (or, for the 10 L2-specific subject-level variables, only over the L2 subjects) and the full set of 58 regular and irregular verbs. Seven of the 26 variables were associated: item order, follows a real (vs. novel) verb, initial phoneme is (or is not) a plosive, initial phoneme is (or is not) a fricative, number of synonym sets, age, and self-consciousness.

All variables that were associated were then entered as fixed main effects into the mixed-effects model described above. In order to evaluate whether any of the item-level covariates should be included as random effects (by subject), which would allow the model to account for differences in the effect of the covariate across subjects, each covariate was separately entered as a random effect into a model which was otherwise identical to the model with the covariates included only as fixed effects. Any variable whose inclusion as a random effect resulted in a large reduction of Akaike’s Information Criterion (AIC, a smaller-is-better measure of model fit), as compared to the model with covariates entered only as fixed effects, was retained as a random effect (as well as a fixed effect) in the final mixed-effects model; otherwise it was included only as a fixed effect. Of the seven variables...
Table 1. The fourteen potentially confounding item-level variables examined for possible inclusion as covariates.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Item order</strong></td>
<td>The number of items presented prior to a given item. Including item order in the model allows one to account for variability attributable to presentation order (e.g., due to practice effects within the task). Order is likely to be more influential for earlier items, with order effects diminishing rapidly as participants become more comfortable with the task; therefore, the natural logarithm of item order was used.</td>
</tr>
<tr>
<td>Follows same inflectional class</td>
<td>Whether or not the previously presented verb was of the same class (i.e., regular or irregular) as the current item. Included because repeating a similar response (-ed-affixed) or producing a different type of response may affect RTs.</td>
</tr>
<tr>
<td>Follows real verb</td>
<td>Whether the previous verb was real or novel. Included because switching from a novel to a real verb may affect processing time.</td>
</tr>
<tr>
<td>Stem and past tense phonological length</td>
<td>The phonological length of each item is likely to affect response time, at the very least because phonological length affects working memory (Caplan, Rochon &amp; Waters, 1992), and the stem is likely to be held in working memory before production of the past tense form. Additionally, since longer spoken forms may require more time for syllabification and articulatory planning than shorter ones (Levelt, Roelofs &amp; Meyer, 1999), the phonological length of the past tense form produced by the participants may also affect response times. Therefore both stem length and past tense length were assessed as possible covariates. In both cases length was computed as the number of phonemes, with diphthongs counted as one phoneme.</td>
</tr>
<tr>
<td>Phonological neighborhood</td>
<td>A measure of the frequency of phonologically similar and dissimilar verbs. Included to account for the influence of verbs with similar or dissimilar stem–past phonological transformations (e.g., the processing of ring–rang may be improved by spring–sprang, but weakened by bring–brought) (for more details, see Prado &amp; Ullman, 2009; Walenski et al., 2007).</td>
</tr>
<tr>
<td>Initial plosive</td>
<td>A binary variable indicating whether the initial sound of the subject’s response was a plosive. Included because this can affect computer-recorded response time measurements (plosives tend to be detected faster than fricatives; Kessler, Treiman &amp; Mullennix, 2002).</td>
</tr>
<tr>
<td>Initial fricative</td>
<td>A binary variable describing whether the initial sound of the subject’s response was a fricative. See just above.</td>
</tr>
<tr>
<td>Synonym sets</td>
<td>An estimate of the number of meanings of a given verb, which can predict whether English verbs are regular or irregular. Following Baayen and Moscoso del Prado Martin (2005), this was computed for each verb as the natural logarithm-transformed number of synonym sets (groupings of words that represent one underlying lexical concept) in which the verb appears, based on the synonym sets in WordNet, an online database of English (Fellbaum, 1998).</td>
</tr>
<tr>
<td>Noun-to-verb ratio</td>
<td>An estimate of the likelihood that a given verb has been converted from a noun or into a noun, which can predict whether English verbs are regular or irregular. Computed as the natural logarithm of the ratio of each stem’s frequency as a noun to that form’s frequency as a verb (Baayen &amp; Moscoso del Prado Martin, 2005).</td>
</tr>
<tr>
<td>Alternation classes</td>
<td>A measure of the number of argument structures associated with each verb, which can predict whether English verbs are regular or irregular. This variable was estimated as the natural logarithm-transformed number of alternation classes (i.e., sets of verbs occurring with the same pattern of arguments and adjuncts) in which that verb occurs (Baayen &amp; Moscoso del Prado Martin, 2005; Levin, 1993).</td>
</tr>
<tr>
<td>Imageability</td>
<td>Uninflected-form imageability ratings for the verbs, based on ratings from 1 (low imageability) to 5 (high imageability) (Prado &amp; Ullman, 2009).</td>
</tr>
<tr>
<td>Consistent voicing</td>
<td>A binary variable indicating whether or not the phonemes in the rime of the past tense form exhibit consistent voicing (e.g., the rime of the regular past tense felled is consistently voiced, as both /l/ and /d/ are voiced, whereas the rime of the irregular past tense felt is not, as /l/ is voiceless). This variable was examined because evidence suggests that two voiced phonemes within a coda may be less perceptually distinct, potentially accounting for differences in performance between regular (always consistently voiced) and irregular (not always consistently voiced) verbs (Bird, Lambon Ralph, Seidenberg, McClelland &amp; Patterson, 2003).</td>
</tr>
<tr>
<td>Phonological changes</td>
<td>The number of phonological changes between the stem and past tense form (computed as in previous studies; Marcus, Pinker, Ullman, Hollander, Rosen &amp; Xu, 1992; Prado &amp; Ullman, 2009). Past tenses of verbs with fewer stem–past changes may potentially be easier to produce.</td>
</tr>
</tbody>
</table>
Table 2. *L1 and L2 biographical variables.*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Age</th>
<th>Years of education</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L1 Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>36</td>
<td>23.9 (8.0)</td>
<td>15.1 (1.9)</td>
</tr>
<tr>
<td>Female</td>
<td>34</td>
<td>26.1 (7.7)</td>
<td>15.7 (1.9)</td>
</tr>
<tr>
<td><em>t</em>-test</td>
<td></td>
<td><em>t</em>(68) = 1.17, <em>p</em> = .246</td>
<td><em>t</em>(68) = 1.31, <em>p</em> = .193</td>
</tr>
<tr>
<td>Total</td>
<td>70</td>
<td>25.0 (7.9)</td>
<td>15.4 (1.9)</td>
</tr>
<tr>
<td><strong>L2 Subjects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>28</td>
<td>35.6 (7.6)</td>
<td>18.9 (1.8)</td>
</tr>
<tr>
<td>Female</td>
<td>28</td>
<td>34.1 (9.6)</td>
<td>17.5 (2.8)</td>
</tr>
<tr>
<td><em>t</em>-test</td>
<td></td>
<td><em>t</em>(54) = 0.65, <em>p</em> = .519</td>
<td><em>t</em>(54) = 2.24, <em>p</em> = .029</td>
</tr>
<tr>
<td>Total</td>
<td>56</td>
<td>34.8 (8.6)</td>
<td>18.2 (2.4)</td>
</tr>
<tr>
<td><strong>L1 vs. L2 <em>t</em>-test</strong></td>
<td></td>
<td><em>t</em>(124) = 6.68, <em>p</em> &lt; .001</td>
<td><em>t</em>(124) = 7.19, <em>p</em> &lt; .001</td>
</tr>
</tbody>
</table>

*Note:* Age and years of education shown in means (with SDs). Although the L1 and L2 participants differed significantly in age and years of education, these variables were considered as covariates in all analyses.

Table 3. *L2-specific biographical variables.*

<table>
<thead>
<tr>
<th></th>
<th>Age of Arrival</th>
<th>Age of first exposure</th>
<th>Length of residence (in years)</th>
<th>Years of formal English instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L2 Subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>28.3 (5.1)</td>
<td>13.9 (7.6)</td>
<td>7.3 (4.0)</td>
<td>6.6 (4.9)</td>
</tr>
<tr>
<td>Female</td>
<td>25.6 (5.7)</td>
<td>12.4 (4.8)</td>
<td>7.4 (5.6)</td>
<td>6.9 (3.5)</td>
</tr>
<tr>
<td><em>t</em>-test</td>
<td><em>t</em>(54) = 1.86, <em>p</em> = .069</td>
<td><em>t</em>(54) = 0.87, <em>p</em> = .391</td>
<td><em>t</em>(54) = 0.03, <em>p</em> = .978</td>
<td><em>t</em>(54) = 0.33, <em>p</em> = .745</td>
</tr>
<tr>
<td>Total</td>
<td>27.0 (5.5)</td>
<td>13.2 (6.3)</td>
<td>7.3 (4.8)</td>
<td>6.7 (4.2)</td>
</tr>
<tr>
<td>Identification with US</td>
<td>(scale of 1 to 5)</td>
<td>Self-consciousness</td>
<td>(scale of 1 to 5)</td>
<td>Desire to stay in US</td>
</tr>
<tr>
<td>Male</td>
<td>2.9 (1.1)</td>
<td>2.7 (1.3)</td>
<td>4.9 (0.3)</td>
<td>3.6 (1.3)</td>
</tr>
<tr>
<td>Female</td>
<td>2.6 (1.1)</td>
<td>2.6 (1.6)</td>
<td>4.9 (0.3)</td>
<td>3.5 (1.2)</td>
</tr>
<tr>
<td><em>t</em>-test</td>
<td><em>t</em>(54) = 0.93, <em>p</em> = .359</td>
<td><em>t</em>(51) = 0.23, <em>p</em> = .820</td>
<td><em>t</em>(54) = 0.23, <em>p</em> = .822</td>
<td><em>t</em>(54) = 0.11, <em>p</em> = .916</td>
</tr>
<tr>
<td>Total</td>
<td>2.7 (1.1)</td>
<td>2.7 (1.4)</td>
<td>4.9 (0.3)</td>
<td>3.5 (1.2)</td>
</tr>
</tbody>
</table>

*Note:* All variables are reported as Means (with SDs). The prompts used for the four attitude and motivation variables were as follows. Identification with US culture: “How strongly would you say you identify with American culture? How would you rate your identification on a scale from 1 to 5, where 1 means not at all and 5 means that you feel like a complete American?” Self-consciousness: “Have you felt self-conscious (i.e., embarrassed) while learning or using English in the U.S.? How would you rate your score on a scale from 1 to 5, where 1 means not at all and 5 means very self-conscious?” Motivation: “How important is it to you to be able to speak English well? How important is it to you to be able to speak English well?” Desire to stay in US: “How strong is your desire to stay in the U.S.? How would you rate your score on a scale from 1 to 5, where 1 means not at all and 5 means very strong?”

that were associated, only item order improved model fit and thus warranted inclusion as a random factor (in addition to inclusion as a fixed effect). Thus, the final mixed-effects model included seven covariates, one of which was included as a random as well as a fixed effect.

Finally, a second model was constructed to specifically examine effects of age of arrival and (ln-transformed) length of residence. This model was identical to the one described above, to maximize comparability, but also included interactions between LoR, native language (English, Spanish, Chinese), verb-type, and frequency, and analogously between AoA, native language, verb-type, and frequency. Both AoA and LoR were held at zero for the L1 subjects; note that this does not influence frequency effects for the L2 subjects, due to the presence of native language in the interaction.
Results

Frequency effects

Before directly comparing the L1 and L2 subject groups, we first wanted to determine whether the native language of the L2 subgroups (Chinese and Spanish) differentially affected the pattern of frequency effects. The interaction between native language of L2 speakers (Chinese vs. Spanish), verb-type (regular vs. irregular) and frequency was not significant ($t(1927) = 1.19, p = .236$), indicating that frequency effects did not show different patterns across regulars and irregulars for the Chinese and Spanish native speakers. Moreover, the magnitude of the frequency effects did not differ between Chinese and Spanish speakers for either regulars or irregulars (regulars: $t(1927) = 1.19, p = .236$; irregulars: $t(1914) = 1.51, p = .131$). Thus, the two subgroups were analyzed as a single L2 group in all further analyses.

We next examined the pattern of frequency effects in L1 and L2 regulars and irregulars. The interaction between group (L1 vs. L2), verb-type (regular vs. irregular) and frequency was significant ($p = .034$), indicating that frequency effects showed different patterns across regulars and irregulars in the L1 and L2 speakers (Figure 1). Whereas for the L1 speakers, the magnitude of frequency effects was significantly greater for irregulars than regulars ($p = .013$), this difference was not significant for the L2 speakers ($p = .196$). Additionally, the L2 speakers showed significantly greater frequency effects than the L1 speakers on regulars ($p < .0001$) but not on irregulars ($p = .168$).

These results do not appear to be due to confounding statistical or experimental factors. First, the reduced frequency effects for L1 regulars are not explained by ceiling effects, lower variability, or lack of power. An examination of ln-transformed response times showed no differences between L1 regulars and irregulars for either
the ln-transformed RTs or their variances (paired t(69) = 0.41, p = .685; for the F-test for equality of variance, F(28,28) = 1.28, p = .523). Moreover, the accuracy of the L1 regulars was the highest of L1 and L2 regulars and irregulars (see Table 4; L1 regulars vs. L1 irregulars: paired t(69) = 9.27, p < .0001; L1 vs. L2 regulars: unpaired t(124) = 10.13, p < .0001). Therefore, since only RTs to correct responses were analyzed, and there were more L1 than L2 subjects, the L1 regulars had more data points than any of the other conditions, and thus the reduced frequency effects are not due to lower power. Further, L2 irregulars, which had the lowest accuracy rate and thus the fewest data points, did indeed show frequency effects. Second, confounding variables did not explain the observed pattern. The regulars and irregulars were matched on stem, lemma, and past tense frequency. Moreover, 26 potentially explanatory variables (14 item-level and 12 subject-level) did not appear to account for the results. Note, as discussed in the Methods above, that only a small subset of these variables was actually included as covariates (based on their association with response time, the dependent variable), greatly reducing the risk of overfitting the data (as compared to a model in which all 26 covariates would have been included). Additionally, no subjects had any neurological, psychiatric, or learning disorders. Finally, note that any subject-related differences between L1 and L2 speakers unrelated to language use would need to explain the specific pattern that was observed – that is, the specific pattern of the group by verb-type interaction for frequency effects.

Although the L1 speakers showed stronger frequency effects for irregulars than regulars, the L1 regulars still elicited significant frequency effects (p = .016; Figure 1). This was in fact expected, given that in a previous study reporting the same set of L1 subjects we had observed such a pattern, with further analyses revealing that frequency effects on regulars were largely explained by a sex difference, with females but not males showing the effects (Prado & Ullman, 2009). We therefore examined frequency effects across L1 and L2, separately for the two sexes. For the male participants (Figure 2), the three-way interaction between group (L1 vs. L2), verb-type (regular vs. irregular), and frequency was significant (p = .015). Irregulars showed frequency effects in both L1 and L2 males, whereas regulars showed frequency effects only in L2 males. Moreover, the magnitude of the effects was significantly greater for L1 irregulars and L2 regulars than for L1 regulars (Figure 2). In contrast, for the female participants (Figure 3), the three-way interaction between group, verb-type and frequency was not significant (p = .535), and both L1 and L2 speakers showed frequency effects for both regulars and irregulars. (Note that the magnitude of the frequency effect for regulars was greater for L2 females than L1 females, consistent with the greater storage of regulars in L2 than L1, even among females; however, since the three-way interaction was highly nonsignificant, examination of this effect is strictly speaking not warranted, and we do not discuss it further.)

Additional analyses showed that among the L1 speakers, the expected (Prado & Ullman, 2009) interaction among sex, verb-type, and frequency approached significance (t(3706) = 1.88, p = .060). Moreover, the finding that regulars showed frequency effects for females (p = .001; Figure 3) but not males (p = .752; Figure 2) was reflected in a significant difference in the magnitude between these effects (t(3712) = 2.25, p < .025), whereas the sexes did not differ in the magnitude of frequency effects on irregulars (t(3749) = 0.38, p = .702). (Note that these results differ slightly from those reported in Prado and Ullman, due to the different regression models used, i.e., the inclusion here of the L2 subjects as well as somewhat different covariates.) In contrast, among the L2 speakers, the interaction of sex, verb-type, and frequency was not significant (t(1925) = 0.58, p = .563), consistent with the finding that both sexes showed frequency effects for both regulars and irregulars (Figures 2 and 3).

These sex-related findings do not appear to be explained by confounding statistical or experimental factors. In particular, the lack of frequency effects for male L1 regulars is not due to ceiling effects, lower variability, or lack of power. There was no significant difference in ln-transformed mean response times or variances between L1 male regulars and irregulars (Table 5; paired t(35) = 0.53, p = .601; F_{var}(28,28) = 1.34, p = .442). In addition, the L1 males produced regulars more accurately than irregulars (Table 5; paired t(35) = 6.48, p < .0001), and not significantly differently from L1 females’ regularly (unpaired t(68) = 1.14, p = .256). Thus, since only RTs to correct responses were analyzed, and there were slightly more L1 males than L1 females (Table 2), the lack of frequency effects for L1 male regulars is not due to lower power. Finally, the observed differences do not seem to

### Table 4. Mean accuracy and response times (and standard errors, SE).

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (SE)</th>
<th>Response Times: raw, ln-transformed (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1 Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulars</td>
<td>99.0% (0.2%)</td>
<td>1581 ms, 7.37 (0.010)</td>
</tr>
<tr>
<td>Irregulars</td>
<td>94.5% (0.5%)</td>
<td>1577 ms, 7.36 (0.009)</td>
</tr>
<tr>
<td>L2 Subjects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulars</td>
<td>74.7% (1.1%)</td>
<td>1837 ms, 7.52 (0.014)</td>
</tr>
<tr>
<td>Irregulars</td>
<td>51.6% (1.2%)</td>
<td>1978 ms, 7.59 (0.017)</td>
</tr>
</tbody>
</table>

Note: Irregulars were significantly less accurate and slower (ln(RTs)) than regulars for both L1 and L2 speakers (p < .005), with the exception of response times for L1 speakers (p = .685). L2 speakers were significantly slower and less accurate than L1 speakers on both regulars and irregulars (p < .01).
be explained by factors unrelated to sex, in particular the various item- and subject-level variables described above.

**Length of residence and age of arrival effects**

To examine whether and how length of residence and age of arrival might influence the storage versus composition of regulars in L2, we constructed a new mixed-effects regression model that was identical to the one used above, but also included LoR and AoA (see Analysis in the Material and methods section above).

First, we examined length of residence in L2, specifically the effect of the natural log of LoR (referred to in these analyses as LoR) on frequency effects while holding AoA constant. The influence of LoR on the magnitude of frequency effects was not significant for either regulars or irregulars, with no difference between them (Figure 4), nor was it significant over both verb-types ($t(1888) = 1.04, p = .301$). Following previous research (see Introduction above) that suggests sex differences in the processing of regulars (but not irregulars), including in L1 in the present study, we then tested for sex differences in the effect of LoR on frequency effects specifically for regulars. LoR significantly influenced frequency effects on regulars in females ($p = .002$), with increasing LoR leading to decreasing frequency effects (Figure 5). The magnitude of this LoR effect was significantly greater than that in males (i.e., there was a three-way interaction between LoR, sex and frequency; $p = .006$), with the males showing a non-significant slope in the opposite direction ($p = .342$). (There were no effects of LoR on irregular frequency effects in either sex, $p_s > .460$.)

To further examine the nature of these sex-related effects, we compared the magnitude of the regular frequency effects to zero, at both the lowest and highest LoRs (lowest over all L2 participants = 3 years, highest = 23 years; note that in the figures these are shown as ln-transformed, consistent with all analyses). This allowed us to test whether each sex did or did not show significant frequency effects for regulars at low and at high LoRs. For females the frequency effect was significantly different from zero at the lowest LoR ($t(1907) = 5.12$, $p < .0001$).

<table>
<thead>
<tr>
<th>Males</th>
<th>Regular Past Tenses</th>
<th>Irregular Past Tenses</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$t(3708) = 3.13$</td>
<td>$t(3824) = 4.51$</td>
</tr>
<tr>
<td></td>
<td>$p = .002$</td>
<td>$p &lt; .0001$</td>
</tr>
<tr>
<td></td>
<td>$t(2768) = 4.24, p &lt; .0001$</td>
<td>$t(2382) = 0.29, p = .771$</td>
</tr>
<tr>
<td></td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$t(1934) = 1.37$</td>
<td>$t(1937) = 1.99$</td>
</tr>
<tr>
<td></td>
<td>$p = .170$</td>
<td>$p = .046$</td>
</tr>
</tbody>
</table>

Figure 2. Males: Past tense frequency effects for regular and irregular verbs, for L1 and L2 speakers.
Storage and composition in adult L2

Figure 3. Females: Past tense frequency effects for regular and irregular verbs, for L1 and L2 speakers.

<table>
<thead>
<tr>
<th>Females</th>
<th>Regular Past Tenses</th>
<th>Irregular Past Tenses</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td><img src="image1" alt="Graph" /></td>
<td><img src="image2" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$B = -0.013$</td>
<td>$B = -0.015$</td>
</tr>
<tr>
<td></td>
<td>$t(3833) = 3.27$</td>
<td>$t(3821) = 3.93$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.001$</td>
<td>$p &lt; 0.0001$</td>
</tr>
<tr>
<td></td>
<td>$t(2770) = 3.19, p = 0.001$</td>
<td>$t(2490) = 0.62, p = 0.535$</td>
</tr>
<tr>
<td></td>
<td><img src="image3" alt="Graph" /></td>
<td><img src="image4" alt="Graph" /></td>
</tr>
<tr>
<td>L2</td>
<td><img src="image5" alt="Graph" /></td>
<td><img src="image6" alt="Graph" /></td>
</tr>
<tr>
<td></td>
<td>$B = -0.041$</td>
<td>$B = -0.034$</td>
</tr>
<tr>
<td></td>
<td>$t(1917) = 0.49$</td>
<td>$t(1970) = 3.06$</td>
</tr>
<tr>
<td></td>
<td>$p = 0.623$</td>
<td>$p = 0.002$</td>
</tr>
</tbody>
</table>

Figure 4. The effect of Length of Residence on frequency effects for regular and irregular verbs, for L2 subjects. In each scatterplot, the line represents the prediction of the mixed-effects model with regard to the effect of ln-transformed LoR on the dependent variable (frequency effect, that is, the slope of the frequency vs. reaction time plot, as in Figure 1), with the effect of all covariates removed. Each point on the graph represents the frequency effect calculated for one L2 speaker.

<table>
<thead>
<tr>
<th>Regular Past Tenses</th>
<th>Irregular Past Tenses</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image7" alt="Graph" /></td>
<td><img src="image8" alt="Graph" /></td>
</tr>
<tr>
<td>$B = 0.015$</td>
<td>$B = 0.005$</td>
</tr>
<tr>
<td>$t(1893) = 1.35$</td>
<td>$t(1895) = 0.31$</td>
</tr>
<tr>
<td>$p = 0.178$</td>
<td>$p = 0.754$</td>
</tr>
</tbody>
</table>
Figure 5. The effect of Length of Residence on frequency effects for regular verbs, in L2 males and females. Each point on the graph represents the frequency effect calculated for one L2 speaker.

Table 5. Mean accuracy and response times (and standard errors, SE) by sex.

<table>
<thead>
<tr>
<th></th>
<th>Accuracy (SE)</th>
<th>Response Times: raw, ln-transformed (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>L1 Subjects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulars</td>
<td>98.8% (0.3%)</td>
<td>1646 ms, 7.41 (0.014)</td>
</tr>
<tr>
<td>Irregulars</td>
<td>93.7% (0.8%)</td>
<td>1635 ms, 7.40 (0.013)</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulars</td>
<td>99.3% (0.3%)</td>
<td>1514 ms, 7.32 (0.015)</td>
</tr>
<tr>
<td>Irregulars</td>
<td>95.3% (0.7%)</td>
<td>1519 ms, 7.33 (0.014)</td>
</tr>
<tr>
<td><strong>L2 Subjects</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Males</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulars</td>
<td>71.0% (1.6%)</td>
<td>1859 ms, 7.53 (0.020)</td>
</tr>
<tr>
<td>Irregulars</td>
<td>50.2% (1.8%)</td>
<td>2042 ms, 7.62 (0.023)</td>
</tr>
<tr>
<td>Females</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Regulars</td>
<td>78.4% (1.4%)</td>
<td>1818 ms, 7.51 (0.020)</td>
</tr>
<tr>
<td>Irregulars</td>
<td>53.0% (1.8%)</td>
<td>1920 ms, 7.56 (0.024)</td>
</tr>
</tbody>
</table>

Note: Males and females did not differ significantly for any comparison (accuracy or ln-transformed RT) for either verb-type (regulars or irregulars) in either group (L1 or L2) (p ≥ .07).

p < .0001) but not at the highest LoR (t(1882) = 0.33, p = .531). Thus, the females show frequency effects at low but not high LoR. Together with the significant slope reported above, this can be taken to suggest that females initially chunk regulars, but with increasing LoR eventually come to rely instead on composition. Males showed a different pattern, with the frequency effect not significantly different from zero (though approaching significance) at the lowest LoR (t(1881) = 0.54, p = .087), but significantly different at the highest LoR (t(1904) = 3.22, p = .021). Thus, the males show significant frequency effects only at high LoR. Together with the lack of a significant slope (see above), this can be taken to suggest that the males rely significantly on chunking only at high LoR, though they do not significantly change their strategy over the course of exposure, from lower to higher LoR levels.

Second, we examined age of arrival, specifically the effect of AoA on frequency effects in L2, while holding LoR constant. The influence of AoA on the magnitude of frequency effects was borderline significant for regulars, with increasing AoA leading to increasing frequency effects (see Figure 6, p = .053). This AoA effect was significantly greater for regulars than for irregulars (i.e., there was a three-way interaction between AoA, verb-type, and frequency; p = .033), which showed no influence at all of AoA on frequency effects (Figure 6). Unlike LoR, the introduction of sex as a factor yielded no differences between males and females regarding the influence of AoA on frequency effects for regulars (interaction between AoA, sex and frequency; t(1902) = 0.42, p = .677). Finally, we compared the magnitude of the frequency effects for regulars to zero, at both the lowest and highest AoAs (17 and 41 years, respectively). These frequency effects differed from zero at the highest AoA (t(1910) = 3.98, p < .0001) but not at the lowest AoA (t(1917) = 0.80, p = .423). Thus, even though all L2 participants had an AoA of 17 years or older, increasing AoA was associated with greater frequency effects for regulars (but not for irregulars), holding length of residence constant. Specifically, holding LoR constant, L2 participants with an AoA of 17 years showed, on the whole, an absence of frequency effects (suggesting they did not depend on stored regular past tense forms), whereas those who arrived around their early 40s showed clear frequency effects (suggesting a reliance on stored regular past tense forms).

These analyses showing effects of LoR and AoA on frequency effects do not appear to be explained by
confounding statistical or experimental factors. In particular, the sex differences in LoR effects are not due to sex differences in LoR itself (males: $M = 2.01$, $SE = 0.09$, females: $M = 1.94$, $SE = 0.11$, $t(54) = 0.47$, $p = .642$; Levene’s test for equality of variances, $F = 1.68$, $p = .200$). Note that because LoR and AoA are subject-level variables, they do not vary between regular and irregular verbs.

**Discussion and conclusions**

This study examined aspects of the computation of adult-learned second language (L2), and how it compares to L1. Specifically, it investigated the storage vs. composition of English past tense inflected forms in L2 and L1, and tested whether and how the storage/composition distinction may be influenced by various factors: linguistic form (idiosyncratic irregulars vs. rule-governed regulars), (dis)similarity between the L2 and the learner’s native language (Chinese or Spanish), sex (male vs. female), amount of exposure (as measured by length of residence), and adult age of acquisition (as measured by age of arrival).

In a past tense production task, we found that L2 learners of English showed stronger full-form (surface) past tense frequency effects than L1 English speakers for regulars, but not for irregulars, which yielded frequency effects of equal magnitude in the two language groups. The pattern of frequency effects did not differ between Chinese and Spanish native speakers learning English as an L2. Sex differences in regulars (but not irregulars) were observed, such that L1 females but not males showed frequency effects for regulars, whereas in L2 both sexes showed robust effects for regulars. Length of Residence (LoR) predicted frequency effects in regulars (but not irregulars), though only in females, who showed frequency effects at low but not high LoR. Adult Age of Arrival (AoA) predicted frequency effects in regulars (but not irregulars), with these effects emerging only at high AoA.

The results do not appear to be due to confounding statistical or experimental factors, such as differences in variability or power. The regular and irregular forms were matched on stem, lemma and past tense frequency. The findings are also not explained by 14 item-level variables (including measures of phonological complexity, phonological neighborhood, and imageability) and 12 subject-level variables (age, education, and 10 L2-specific measures: age of first exposure, age of arrival, length of residence, and several measures of motivation and the amount of formal English instruction). All of these variables were tested for inclusion in the regression model, and were included only if statistically warranted, thereby avoiding overfitting. Finally, the sensitivity of mixed-effects regression models minimizes the likelihood that null effects (e.g., of regulars in L1) are false negatives, particularly in the context of significant frequency effects in matched control conditions (e.g., irregulars in L1, regulars in L2).

The results are not predicted by single-mechanism models. These models, which posit that all inflected forms are represented and processed in a distributed associative memory, reject the notion of separate compositional processes for inflected forms (Joanisse & Seidenberg, 1999; McClelland & Patterson, 2002; Plunkett & Marchman, 1993; Rumelhart & McClelland, 1986; Woollams, Joanisse & Patterson, 2009). Whereas such models share the assumption made here that the presence of frequency effects suggests storage, they interpret the absence of such effects differently. Because many regulars have similar stem-past phonological transformations, the general phonological pattern across them is expected to be learned in associative memory, which can leave little or
no influence to individual word frequencies (Daugherty & Seidenberg, 1992; Seidenberg, 1992). Thus, regulars may be expected to show phonological neighborhood effects, but not necessarily individual word frequency effects, though such effects could also be observed. The pattern of results reported here does not seem to be expected by such models. Phonological neighborhood did not show any association with response times; that is, there was no evidence of phonological neighborhood effects. Indeed, response times were also not associated with several other measures of phonological complexity, including stem and past tense phonological length, whether or not there was consistent voicing in the past tense form, and the number of phonological changes between the stem and past tense form. Moreover, the only condition in which there was an absence of frequency effects, namely on regulars for males but not females in L1, was predicted on a view that assumes that regulars can be composed (a “dual-system” view) (Prado & Ullman, 2009; Ullman, 2004, 2005; Ullman et al., 2008), whereas we are aware of no such predictions from single-mechanism models. More generally, the pattern of frequency effects observed here, across L1 and L2, over males and females, was largely predicted on a dual-system view, but is not expected by any single-mechanism model we are aware of. Thus, although we cannot by any means rule out the possibility that future single-mechanism models or simulations could explain the findings presented here, this is not yet the case.

Rather, the finding that irregulars showed frequency effects of equivalent magnitude in L1 and L2 speakers, but that there were stronger frequency effects for regulars in L2 than L1, suggests the following. First, consistent with previous studies (see Introduction), irregular past tense forms are always stored, whether in L1 or L2. Second, regulars can be either stored (chunked) or composed, and, crucially, they depend more on storage in L2 than in L1. This pattern for regulars is broadly consistent with previous studies, which generally report a lack of frequency effects for regulars in L1, but show more of a mixed pattern in L2. The present study strengthens the view that, holding other factors constant (including LoR and AoA), at least English regular past tense forms do indeed depend more on storage and less on composition in L2 than in L1.

The finding that Chinese and Spanish native speakers did not show different frequency effect patterns is consistent with at least some previous findings suggesting that the native language of L2 learners does not significantly affect the storage vs. computation of inflected forms in the L2 (Birdsong & Flege, 2001; Silva, 2009; Silva & Clahsen, 2008). In particular, the fact that inflectional morphology is virtually absent in Chinese, but shows complex suffix-based patterns in Spanish, strengthens the view that the structural similarity of a learner’s native language to the L2 does not substantially influence the storage/composition of inflected forms in the L2. However, some caution is warranted regarding this conclusion, since other studies are consistent with the view that L2 learners’ native language might in fact influence the computation of inflected forms (Basnight-Brown et al., 2007; Feldman et al., 2010; Portin et al., 2008). Further studies may clarify this issue.

The observation that sex influenced the storage vs. composition of regulars in L1 but not L2 (over all subjects, i.e., independently of LoR), while not affecting the storage of irregulars, suggests the following. It further emphasizes that irregulars are always stored, in all circumstances by all individuals. As previously reported by Prado and Ullman (2009), it supports the view that in L1, females but not males tend to depend on stored (chunked) representations of regulars, at least in English past tense production. Yet it suggests that in L2 the situation is rather different. Specifically, at least when holding other factors constant (e.g., LoR, AoA, and other variables), L2 learners tend to depend on stored regulars whether they are male or female. In other words, being an L2 learner seems to be more important than sex in determining whether one depends on the storage or composition of regularly inflected forms.

We have seen that LoR predicts frequency effects – but only for regulars, and only in females. This finding, which is the first direct examination of the influence of length of exposure on the storage/composition of inflected forms, suggests the following. Holding AoA and other factors constant, the amount of exposure to the L2 does indeed affect the storage vs. composition of inflected forms, but only for regulars, and apparently only in females. The pattern indicates that women learning an L2 tend to chunk regulars in the first few years of learning (at low LoR), but gradually come to depend on composition with increasing exposure to the language. It is not clear why at high LoR the magnitude of the frequency effect for regulars in females did not differ from zero. This finding suggests that, unlike L1 females, these L2 learners depended on composition rather than on storage. This seems surprising, given that L2 learners are not expected to depend more on composition than L1 speakers – even L2 learners with substantial exposure to the language (Ullman, 2001, 2005). The pattern does not seem to be explained by greater amounts of exposure to English by the L2 than L1 participants (which could lead to a greater opportunity to proceduralize), since the L1 females had on average more exposure (mean 26 years of age) than the L2 females (23 years for the longest LoR). Another possibility is that in fact the L2 females did not rely on composition at high LoR, and that the finding here is a statistical fluke, for example due to linear modeling. Then again, it is also possible that this is not an anomalous result, and that indeed adult female L2 learners might eventually depend on composition more than female L1 speakers. Further studies examining this issue seem warranted. In contrast, the data suggest that
LoR does not significantly affect the storage/composition of regulars in men, who show a different pattern than women: no significant frequency effects for regulars at low LoR (consistent with at least some composition), and significant frequency effects at high LoR (suggesting chunking). However, the lack of a significant effect of LoR on frequency effects for regulars in men suggest that this pattern must be treated with caution, and further studies are needed. The storage of irregulars, on the other hand, does not show any evidence of being influenced by LoR, further supporting the view that irregulars always depend on memorized representations.

The finding that adult age of arrival (AoA) predicts frequency effects for regulars, but not irregulars, holding LoR and other factors constant, is also a novel result. Previous L2 studies have not examined the influence of AoA (or other measures of age of acquisition) on the storage vs. composition of inflected forms. The observed pattern suggests that, **EVEN DURING ADULTHOOD**, age of acquisition – at least when measured with age of arrival, in which subsequent exposure is largely naturalistic – has an effect on the computational mechanisms underlying inflected forms. Specifically, later **ADULT AoA** leads to an increased reliance of regulars on storage and less on composition. Holding other factors constant, individuals who had an AoA in early adulthood (around age 17) tend to compose regulars, whereas individuals with later AoAs, in particular around age 40, tend to chunk them. In contrast, AoA does not appear to have any effect on irregulars, which are always stored.

Overall, these results, together with previous findings, strengthen our theoretical understanding of the storage vs. composition of inflected forms in L2 and L1. In particular, the data support the perspective endorsed by Clahsen and Ullman, as follows. Irregularly inflected forms do indeed always seem to be stored in lexical memory, whether in L1 or L2. Regularly inflected forms, and perhaps other rule-governed complex linguistic forms, are generally composed in L1, but depend more on storage and lexical memory in L2. This dependence, however, changes as a function of exposure (or proficiency), with increasing levels associated with an increasing reliance on composition. The present study is the first to show such a change in reliance (in females), specifically as a function of length of residence.

This study also further strengthens and extends the views of Ullman and Clahsen in other ways. First of all, the sex differences are consistent with and further specify Ullman’s declarative/procedural model of language (Prado & Ullman, 2009; Ullman, 2001, 2004, 2005; Ullman et al., 2008). The finding that L2 females but not L2 males seem to rely significantly on chunking for regulars at low LoR is consistent with the view that females are better at declarative memory, and possibly worse at procedural memory (Introduction), since the females but not the males initially rely heavily on storage. With increasing LoR, the data suggest that the L2 females gradually proceduralize the rule-governed pattern in regulars, while the males either show no change over time (consistent with the lack of a significant influence of LoR on frequency effects for regulars in males), or eventually depend on storage (consistent with the significant frequency effects for regulars at high LoR in males). Overall, these sex differences in the effect of LoR on frequency effects for regulars, as well as in the pattern of frequency effects in L1, further support the importance of sex as a factor that affects storage vs. composition, in both L2 and L1. The results strengthen the view that sex should be included as an experimental factor in studies of language (Prado & Ullman, 2009; Ullman, 2004; Ullman et al., 2008).

The lack of any significant influence of the L2 learner’s native language on frequency effects is consistent with at least some previous studies, and further strengthens Clahsen’s view that L1–L2 structural similarity does not appear to play an important role in the storage/composition of inflected forms (Introduction). However, it is important to emphasize that other research, including from the field Second Language Acquisition, does support L1–L2 transfer effects (see Clahsen et al., 2010; Tolentino & Tokowicz, 2011), suggesting that this issue is still far from resolved.

The finding that adult AoA affected the storage vs. composition of regular past tenses was not predicted by either Clahsen or Ullman, but is also not inconsistent with either view. The observed pattern suggests that the effect of age of acquisition on L2 is not in fact limited to the period before adulthood, as has been suggested (Johnson & Newport, 1989; Lenneberg, 1967). Rather, contrary to strong versions of the critical period hypotheses, it appears that the computational mechanisms underlying at least some aspects of language (rule-governed aspects of inflectional morphology) continue to be affected by age of acquisition – at least by age of arrival – well into adulthood. Similarly, the confirmation of Ullman and Clahsen’s predictions that increasing exposure or proficiency can lead to native-like grammatical mechanisms (composition), even in adulthood, seems problematic for strict views of the critical period hypothesis.

Whereas the findings in this study were either predicted by or compatible with the theories proposed by Ullman and Clahsen, they do not appear to be consistent with the other theoretical perspectives discussed above (see Introduction). The finding that LoR influences the mechanisms underlying regulars, but not irregulars, does not seem to be expected by Paradis. The divergence of computational mechanisms, i.e., storage/composition, between L1 and L2 in some circumstances, but not others, in a manner that is modulated by particular item-level
(e.g., regular vs. irregular) and subject-level (e.g., sex, LoR, AoA) factors, is problematic both for theories that claim that L2 always depends on L1 mechanisms, and for theories that claim that L1 and L2 always depend on different mechanisms (Introduction).

This study has various limitations that may be addressed by future research. First, we focused here on LoR, and did not attempt to measure proficiency. We took this approach because LoR seems to be a more objective and replicable measure than proficiency measures. Moreover, previous studies suggest that LoR might be a better predictor than proficiency of changes in the storage/composition of regulars. Finally, LoR and proficiency measures are, nevertheless, usually highly collinear, and thus difficult to tease apart. However, focusing only on LoR is not a perfect approach. For example, length of residence might not correlate highly with proficiency in closed immigrant communities, in which some of our subjects might have been living. Therefore, it would be informative if future studies were to examine the role of both LoR and proficiency (e.g., using a cloze test, see Tremblay, 2011) in the storage and computation of inflected forms. Second, and along the same lines, the present study only tested the effects of one measure of the amount of exposure, that is, LoR, and one measure of the age of acquisition, that is, age of arrival (AoA). We focused on these measures because they allowed us to examine the effects of exposure and age of acquisition in largely naturalistic, immersion-based, learning contexts. However, it is possible that other measures of the amount of exposure and age of acquisition might lead to different patterns of influence on storage/composition – particularly since recent studies suggest that the learning context (e.g., immersion vs. classroom) may indeed affect the neurocognition of L2, with immersion being more likely than classroom contexts to lead to L1-like neurocognition for syntax (Morgan-Short, Finger, Grey & Ullman, 2012; Morgan-Short, Sanz, Steinhauer & Ullman, 2010; Morgan-Short, Steinhauer, Sanz & Ullman, 2012). Third, even though the number of subjects in the present study was not small, the relatively large number of factors included in the final mixed-effects regression models (seven covariates, in addition to several independent measures) suggests the desirability of replicating the study, ideally with a larger number of subjects. Additionally, future studies might test for the inclusion of covariates in more sophisticated ways, including the examination of interactions (e.g., an item-level covariate might be associated with reaction times for irregulars but not for regulars). Finally, the present study focused solely on inflectional morphology, specifically on English past tense in a production task. Further research is needed to elucidate whether the findings reported here might generalize to other tasks (Woollams et al., 2009), other inflectional paradigms (in English and other languages), and other aspects of language, including derivational morphology and syntax.

In sum, this study examined the storage vs. composition of regular and irregular English past tense forms in adult-learned L2 as compared to L1. The evidence suggests that irregular forms are always stored, in both L1 and L2. In contrast, English regular past tense forms can either be composed or stored, as a function of multiple factors, some of which interact. These include the sex of L1 and L2 speakers, and both the length of residence and adult age of arrival of L2 learners. In particular, both sexes seem to generally store regulars in L2, but only females do so in L1; increasing LoRs in L2 learners lead to less dependence on storage and more on composition, but only in females; and a later AoA during adulthood leads to less reliance on composition and more on storage. In contrast, the findings in this study suggest that the native language of L2 learners might not affect the storage/composition of inflected forms. Overall, the data presented here, together with previous evidence, suggest that inflected forms in L2 do not always depend on the same mechanisms as in L1 (contra Ellis, 2003; Indefrey, 2006), nor do they always depend on different mechanisms than in L1 (contra Bley-Vroman, 1989). Rather, inflected forms in L2, and thus at least to some extent linguistic forms more generally, depend either on the same or on different mechanisms as in L1, and, crucially, this dependence varies as a function of multiple item- and subject-level factors (Clahsen & Felser, 2006b; Ullman, 2001, 2005, 2012).

References


Bird, H., Lambon Ralph, M. A., Seidenberg, M. S., McClelland, J. L., & Patterson, K. (2003). Deficits in phonology and


Church, K. (1988). A stochastic parts program and noun phrase parser for unrestricted text. Presented at the Second Conference on Applied Natural Language Processing, Austin, TX.


