Late L2ers can acquire grammatical features that do not occur in their L1: Evidence from the effect of animacy on verb agreement in L1 Chinese

Henrietta Lempert[1]

Abstract Second language (L2) learners often have persistent difficulty with agreement between the number of the subject and the number of the verb. This study tested whether deviant L2 verb number agreement reflects maturational constraints on acquiring new grammatical features or resource limitations that impede access to the representations of L2 grammatical features. L1-Chinese undergraduate students at three age of arrival (AoA) levels were tested for online verb agreement accuracy by completing preambles in three animacy combinations: animate–inanimate [AI; e.g., *The officer(s) from the station(s)*], inanimate–animate [IA; e.g., *The letters from the lawyer(s)*], and inanimate–inanimate [II; e.g., *The poster(s) from the museum(s)*]. AI should be less costly to process than IA or II sequences, because animacy supports the subject in AI but competes with the subject for control of agreement in IA sequences, and is neutralized in II. Agreement accuracy was greater overall for AI than for IA or II, and although an AoA-related increase in erroneous agreement after plural subjects occurred for IA and II, there were no AoA effects for AI. Higher scores on memory tasks were associated with greater agreement accuracy, and the memory tasks significantly predicted variance in erroneous agreement when AoA was partialed out. The fact that even late learners can do verb agreement in the case of AI demonstrates that they can acquire new grammatical features. The greater difficulty with agreement in the case of IA or II than of AI, in conjunction with the results for the memory tasks, supports the resource limitation hypothesis.

In many languages, verbs overtly agree with the subject in number, person, and/or gender. As compared to languages such as French and Italian, English has an impoverished system of verb agreement: Apart from auxiliary and copular verbs such as *be* and *have*, singular and plural verb forms are morphologically identical in the present tense, with the sole exception of third person singular, in which a phonological variant of /s/ attaches to the verb stem. Yet despite the overt simplicity of the English verb agreement paradigm, first language (L1) Chinese learners of English as a second language (L2ers) are notoriously unreliable in supplying verbs with inflections for agreement and tense, even after many years of immersion in the L2 (e.g., Lardière, 2003, 2006). For some language researchers, this inflectional difficulty is in line with the view that the acquisition of L2 morphology and syntax is constrained by a critical period that precludes attaining native proficiency in grammatical features present in the L2 but absent in the L1 (e.g., Hawkins, 2000, 2001; Hawkins & Chan, 1997; Hawkins & Liszka, 2003). However, other researchers attribute grammatical deviance in even proficient L2ers to resource limitations rather than to deficient mental representations for the L2 grammatical features (e.g., McDonald, 2006; Miyake & Friedman, 1998; Prévost & White, 2000). The present work is an examination of whether online difficulty with verb number agreement in L1-Chinese L2ers reflects biological constraints on acquiring grammatical features that are not instantiated in the L1, or

✉ Henrietta Lempert
lempert@psych.utoronto.ca

[1] Department of Psychology, University of Toronto, Toronto, Ontario, Canada

whether late L2ers may possess these features, but cannot always gain access to them when the task demands are high.

The critical or sensitive period hypothesis, as originally formulated by Lenneberg (1967) for L1 acquisition, and by Penfield and Roberts (1959) for L2 acquisition, assumes that an increasing decline in brain plasticity from early childhood to puberty circumscribes the optimal period for native-like attainment of a language. This hypothesis is now widely accepted for L1 acquisition and for L2 phonology (e.g., Flege, Yeni-Komshian, & Liu, 1999; Granena & Long, 2013), but it remains controversial for the acquisition of L2 morphology and syntax. Behavioral evidence for the hypothesis largely consists of an inverse correlation between accuracy on grammaticality judgment tests and age of arrival (AoA) in the L2-dominant country; the closure of the critical period is defined as the age at which the association is no longer significant (Johnson & Newport, 1989). However, the limits of the purported critical period vary more with the similarity of the L1 and L2 grammars than with age. For L1 Chinese and Korean learners of L2 English or Spanish, peak sensitivity for L2 acquisition is seen until about age 7, followed by an age-related decline that ends somewhere between age 15 (Granena & Long, 2013) and age 23 (Flege et al., 1999). For L1 Spanish learners of L2 English, native-like attainment is possible until about age 17 and is followed by a gradual age-related decline that extends to age 44 (Birdsong & Molis, 2001). Moreover, AoA is often confounded with variables that also influence end-state proficiency such as years of exposure to the L2: When Flege et al. held AoA constant in L1 Korean groups that differed in their numbers of years of education in the United States, L2 grammatical proficiency depended on years of L2 schooling.

Representational deficiency or performance limitations?

For some language researchers, morphosyntactic deficiency in advanced late-L2 learners reflects performance limitations rather than defective or absent mental representations for the grammatical features in question. Prévost and White (2000; White, 2010) attributed the omission of obligatory inflections on English verbs (i.e., past tense -*ed* and third person singular -*s*) to temporary difficulty accessing representations for the inflections during communication pressure. McDonald (2006) ascribed grammatical deficiency to limitations of L2 resources, and reported a significant correlation between reaction times on an auditory grammaticality judgment task and verbal working memory in late L2ers from a variety of L1s. She further demonstrated that natives responded to the grammaticality test like L2ers under conditions intended to stress processing resources (white noise, memory load, and compressed speech). Hopp (2010, Exp. 4) demonstrated that the processing of morphosyntax in German sentences presented word by word deteriorates in natives at elevated processing speed. Native German speakers became less accurate at detecting violations of case and gender marking with successive reductions in the rate at which words were presented; at the slowest rate (250 ms), natives surpassed proficient L1 Dutch and English in accuracy, but at a rate of 71 ms, their accuracy deteriorated to the level seen in L1 Dutch and English at 250 ms.

One might argue that interventions such as white noise and speeded word presentation impact grammatical judgments by reducing the accurate perception of grammatical features rather than by degrading processing of the features. However, this possibility does not seem to apply to the concurrent memory load paradigm. There is convincing evidence from French natives that extrinsic memory load adversely affects subject–verb agreement accuracy (Fayol, Hupet, & Largy, 1999; Fayol, Largy, & Lemaire, 1994; Hupet, Fayol, & Schelstraete, 1998), and evidence that extraneous memory load has particularly severe consequences for agreement accuracy in individuals with low working memory span (Hartsuiker & Barkhuysen, 2006). It thus seems as if access to (or retrieval of) mental representations for grammatical features is delayed or inefficient when processing resources are strained, even in the case of native speakers.

In summary, there is support for maturational constraints on ultimate proficiency in L2 grammars, but there is also support for a resource limitation account, although the evidence has come largely from research with native speakers rather than late L2ers.

The challenge of inflections for L1 Chinese

The Chinese language is ideally suited for examining whether L2ers can acquire grammatical features that are not part of their L1 experience, and thus cannot be acquired through transfer from the L1. From a critical-period perspective, post-pubescent L1-Chinese L2ers should have difficulty mastering inflectional morphology such as subject–verb agreement, tense, or number, because Chinese does not use inflections or otherwise mark agreement. Grammatical and semantic relations are conveyed through context, word order, animacy, and free standing function-like words. Verbs can occur with particles that carry aspectual and modal information, but verbs do not attach inflections for person or number. Also, common nouns do not bear obligatory inflections for number, gender, or case. For example, a speaker of Mandarin Chinese might say *san zhi bi* ("three stick pen") where an English speaker would say *three pens*; the number word *san* ("three") is followed by the classifier *zhi* ("stick"), but the noun (*bi*) is the same for singular and for multiple instances of the referent (Cheng & Sybesma, 1999; C. N. Li & Thompson, 1989).

Near-native attainment of L2 inflectional morphology is possible in the case of grammatical features that also occur in the L1 (Hopp, 2010). In contrast, English verbal inflectional morphemes of tense and agreement are problematic even for highly proficient L1-Chinese L2ers. Lardière (1998a, 1998b, 2003, 2006) examined morphological variability in the speech of Patty, an L1 speaker of Mandarin and Hokkien who moved to the United States at the age of 22. After 10 years' residence in the United States, Patty's production of obligatory inflections for past tense -*ed* and third person singular -*s* was extremely deficient (34 % and 5 % production, respectively); 8 years later, there was no improvement. Similarly, Goad, White, and Steele (2003) reported low production of tense and agreement morphology in L1 Mandarin Chinese late learners' descriptions of pictures (length of residence in Canada = 6 months to 5 years).

Although the above reports are consistent with the tenets of the critical-period hypothesis, there is also evidence that late L2ers are able to acquire representations for tense and agreement. Goad et al. (2003) reported native-like performance when their sample of inflection-deficient L2ers was tested for knowledge of tense and agreement morphology through grammaticality judgments rather than production. Similarly, L1-Mandarin late L2ers supplied past and perfective inflections 90 % of the time when producing sentences they had previously read under conditions directing attention to inflections (Goad & White, 2008).

Copular *be* and auxiliary *be*, *have*, and *do* can be used to assess underlying knowledge of tense and agreement because they do not add inflections for person, number, or tense. High production (about 90 %) of *be* variants (*is*, *was*) and of related verbal forms (*has*, *have*) does in fact occur in the spoken and written English of L1 Chinese (Goad et al., 2003; Lardière, 1998a, 1998b, 2003, 2006). On the basis of such findings, it has been proposed that late learners possess intact mental representations for tense and agreement, but may be unable to access these representations and/or procedures for integrating inflections with syntactic features when resources for processing the L2 are stressed (Prévost & White, 2000; White, 2010; see also Hopp, 2010).

Empirical studies on verb number agreement

Knowledge of subject–verb agreement in L1-Chinese L2ers has largely been examined through sentence processing rather than production. The paradigm for this work was introduced by Pearlmutter, Garnsey, and Bock (1999, Exp. 1), who studied the trajectory of sensitivity to agreement violations in

natives with self-paced word-by-word reading of grammatical and ungrammatical versions of sentences such as (1) and (2):

(1) The key to the cabinet was/were rusty from many years of disuse. (SS)

(2) The key to the cabinets was/were rusty from many years of disuse. (SP)

In (1), the head noun (*key*) and local noun (*cabinet*) are both singular (SS), whereas in (2), a singular head noun combines with a plural local noun (SP). An effect of grammaticality emerged at the adjective (*rusty*), in the shape of prolonged processing time following agreement violations: Reading times at the verb and adjective were faster for the match (SS) than for mismatch (SP) version, mirroring the attraction effect seen in oral completion of SP preambles such as *The key to the cabinets* (Bock & Miller, 1991). (*Attraction* refers to erroneous agreement with the local noun phrase rather than the subject in mismatch preambles; e.g., *The key to the cabinets were rusty.*)

Jiang (2004, Exp. 1) used self-paced reading to examine online sensitivity to agreement violations in 30 L1-Chinese graduate students who had resided in the United States from 6 months to 5 years and had studied English for a minimum of 7 years, with a start age ranging from 7 to 14. The grammatical versions of SP and SS sentences such as (1) and (2) were presented in the guise of a reading comprehension task. In L1-English controls, reading times were significantly prolonged at the verb and adjective for SP relative to SS (consistent with Pearlmutter et al., 1999). Longer reading times at the verb for SP than for SS sentences were also seen in the L2ers, but the difference was not reliable, and no difference emerged between SP and SS sentences at the adjective. According to Jiang, the absence of attraction reflects insensitivity to the plural inflection on the local noun of SP sentences during online processing. But if the L2ers were insensitive to plural morphology, reading times for plural and singular nouns should be comparable. However, the fact that they took significantly longer for plural than for singular local nouns suggests greater difficulty processing plural than singular nouns, rather than an insensitivity to plural inflections. This difficulty could, in turn, reduce the resources available for detecting deviant verb agreement, especially if L2 resources were limited to begin with.

Jiang (2004, Exp. 2) next examined responsiveness to deviant agreement in sentences with plural verbs and singular local nouns. Grammaticality was manipulated by means of head number, such that sentences with a plural head noun and singular local noun (PS) were grammatical, whereas SS sequences were ungrammatical, as in (3) and (4);

(3) The bridges to the island were about ten miles away. (PS)

(4) *The bridge to the island were about ten miles away. (SS)

In natives, reading times at the postverb phrase were significantly prolonged for ungrammatical sequences, indicating responsiveness to violations of agreement. Longer postverb reading times for ungrammatical SS than for grammatical PS sentences also occurred in the L2ers, but the difference was not significant.

Jiang recognized that the results for the L2ers could also reflect a deficiency in processing subject–verb agreement as well as insensitivity to inflections for plurality. To disambiguate the two possibilities, the participants in Experiment 2 also responded to violations of pronoun-*be* agreement (e.g., *She am a professor.*), and these sentences did elicit reliably longer reading times from the L2ers than did grammatical pronoun-*be* agreement (*I am a professor.*). Jiang concluded that the L2ers had acquired number agreement, but could not distinguish grammatical PS from ungrammatical SS sequences due to insensitivity to the plural morpheme on the head noun of PS sentences. In other words, Jiang seems to argue that the grammatical PS sequences were processed as SS sequences, so that both PS and SS were perceived to contain verb number errors.

It would not be surprising if *be* agreement emerges earlier for pronouns than for nouns, simply because pronoun–*be* concord involves a small set of pairings that are either unique (*I am*) or highly consistent (*s/he/it is/was*, *you/we/they are/were*), and accordingly can be acquired as memorized, unanalyzed chunks. One should thus be cautious about accepting that the L2ers had mastered number agreement due to their sensitivity to deviant pronoun–*be* agreement. In addition, it is not clear whether they were insensitive to inflections for plurality, were insensitive to deviant agreement (implying a representational deficit for agreement), or took longer to access/retrieve the representations for inflections and/or verb agreement online.

The present work used online completion of preambles such as *The letter(s) from the lawyer(s)* to examine the production of subject–verb agreement in L1 Chinese at three AoA levels: early (AoAs = 0 to 10 years old), intermediate (AoAs = 11 to 15), and late (AoAs = 16 to 24). AoA was used to index age at immersion in the L2 and was manipulated in order to examine whether agreement accuracy declines as a function of increased AoA (consistent with the critical-period hypothesis). However, a decline could also result from constraints on L2 processing resources that interfere with efficient access to representations of agreement and procedures for computing agreement. These possibilities could be disambiguated by manipulating preamble difficulty (so that some were easier to process than others). Results indicating that late learners can compute subject–verb agreement (even if this were restricted to easy-to-process preambles) would argue against maturational constraints on acquiring new grammatical features. Evidence that they can compute agreement on the easy but not on the difficult-to-process preambles would be consistent with the resource limitation hypothesis.

The resource limitation hypothesis was also directly evaluated by testing participants on the following memory tasks; digit span forward (DF), digit span backward (DB), and an alphabet-ordering task that required mentally rearranging the order of orally presented words alphabetically (modeled after Craik, 1986). The DF and DB tasks were used because they activate a functional neural system associated with working memory (right dorsolateral prefrontal cortex and bilateral inferior parietal area, although DB activates a larger area including left dorsolateral prefrontal cortex and Broca's area (Gerton et al., 2004)). An inverse relation between scores on the memory tasks and verb agreement errors would strengthen the resource limitation hypothesis, as it would indicate that the likelihood of erroneous agreement was greater for individuals with fewer resources than for those with greater resources (cf. Hartsuiker & Barkhuysen, 2006).

Preamble processing cost was initially varied by manipulating plausibility (e.g., *The cheque from the client . . . void/ dishonest*), but implausible preambles elicited an excessive number of agreement errors even from natives. It was decided to instead manipulate preamble difficulty through animacy, for reasons elaborated below. Processing costs should be lower when animacy relations in the preambles concur with processing expectations and, conversely, higher when the animacy relations conflict with processing expectations.

Manipulating processing cost through animacy

Animates are better remembered than inanimates (Bonin, Gelin, & Bugaiska, 2014). The distinction between animates and inanimates is fundamental in human cognition (Rakison & Poulin-Dubois, 2001) and is incorporated in the agreement systems of many languages (Corbett, 2000, 2006). Animates enjoy a privileged processing status over inanimates in many languages, including English and Chinese, which share a preference for sentences with an animate–inanimate order (as opposed to inanimate–animate order). In English natives, the animate–inanimate order preference can override the default preference for the active voice and motivate production of passives such as (5) below, which has an active counterpart with inanimate–animate order. Similarly, the animacy preference counteracts production of inanimate–animate passives such as (6), and promotes production of the active voice instead:

(5) The nurse was knocked down by the truck.
(6) The cake was decorated by the chef.

Sentences that are incompatible with animacy-processing preferences decrease memory for sentence voice (e.g., Dewart, 1979), override the effectiveness of syntactic priming (Bock, 1986, Exp. 1), and negatively affect preschoolers'

acquisition of the passive voice, relative to compatible sentences (Lempert 1989, 1990).

Animacy is a more crucial cue to grammatical and semantic relations in Chinese than is English, because grammatical relations are not marked through inflections and subject–verb agreement (Matthews & Yip, 2011; Xuefang, 2010). Additionally, word order is more flexible in Chinese relative to English, and thus is not always a good cue to the relations between words in sentences. Wang, Schlesewsky, Bickel, and Bornkessel-Schlesewsky (2009, Exp. 2) reported faster and more accurate processing of Chinese transitive sentences in animate–inanimate order (7) than in sentences in the reverse order (8):

(7)   Yanyuan lijie-le xiaoshuo.
         Actor understand-ASP novel.
(8)   Xiaoshuo jiayu-le yanyuan.
         Novel educate-ASP actor.

Sentence interpretation studies have shown that L1 Chinese with low or intermediate English proficiency rely heavily on animacy information when struggling to interpret sequences such *The monkey the apple bumps*, *The apple bumps the monkey*, and *Bumps the monkey the apple*. With increased proficiency in English, they shift to the native English strategy of assigning the subject to the noun immediately preceding the verb (Liu, Bates, & Li, 1992; Su, 2001a, 2001b).

The present work is based on preambles in three animacy combinations: animate–inanimate (AI), inanimate–animate (IA), and inanimate–inanimate (II). In AI preambles such as (9) below, animacy and word order are congruent cues to the agreement controller (the head noun or subject), whereas in IA preambles such as (10), the processing priority given to animates competes with assigning agreement control to the head noun. In the case of inanimate–inanimate preambles such as (11), animacy is neutralized.

(9)   The chef for the dinner (AI)
(10)   The penalty for the student (IA)
(11)   The piano for the wedding (II)

AI is consistent with animacy preferences in both Chinese and English sentences, and thus allows for transfer of semantic processing strategies from Chinese to English. Accordingly, AI sequences are expected to be easier to process than IA sequences, in which animacy processing priorities compete with assigning agreement control to the head noun. Resolving this conflict is expected to require deeper, more intensive semantic–syntactic processing than in the case of AI preambles, and to diminish the resources needed for computing subject–verb agreement. II preambles should also be more difficult than AI, because word order is the only cue to the subject, but II should be less difficult than IA, due to the

absence of animacy-induced conflict. Deviant agreement should therefore be more common for IA than for AI preambles, with II falling in between.

An animacy-neutralized AA (both animate) condition was not added because representations with overlapping semantic features can create concept confusability in memory (Solomon & Pearlmutter, 2004) and increase aberrant agreement (cf. Barker, Nicol, & Garrett, 2001). There is considerable semantic overlap between noun pairs such as *chef* and *emperor*, because both are living, human, sentient beings, whereas one can choose inanimate noun pairs that share few semantic features (e.g., *design–airplane*).

The hypotheses and predictions can be summarized as follows:

1.   Maturational limits on acquiring new grammatical features will be reflected by an AoA-related increase in deviant verb agreement. Conversely, the absence of AoA differences in verb agreement accuracy would argue against this hypothesis, even if an AoA effect is only absent in the case of AI.
2.   The resource limitation account of agreement errors would be supported by

    (a)   an AoA-related increase in errors on IA and II preambles, concomitantly with its absence for AI, and
    (b)   an inverse correlation between the memory tasks and verb agreement errors.

Method

Participants

A group of 74 Introductory Psychology students at the University of Toronto participated in exchange for course credit; 55 students from 17 to 58 years of age ($M = 21.1$, $SD = 6.5$) claimed Chinese as their first language, and 19 from 17 to 19 years of age ($M = 18.5$) were native English speakers. The L1-Chinese participants were split into three AoA groups according to their stated AoA in an English-dominant country: early ($n = 19$), 0 to 10 years ($M = 4.7$); intermediate ($n = 19$), 11 to 15 years ($M = 13.1$); and late ($n = 17$), 16 to 25 years ($M = 19.0$). The early arrivals included Canadian-born participants who acquired Chinese before English. The accumulated lengths of residence in an English-dominant country ranged as follows: early, 9 to 23 years; intermediate, 1 to 9 years; and late, 6 months to 33 years (see Table 1 for the means and demographic characteristics of the AoA groups). Late arrivals had received at least 4 years' instruction in English prior to arrival and had met the University of Toronto requirements for

Table 1 Demographic characteristics of the early, intermediate, and late age-of-arrival groups: Means (*SD*s) and *F* values

Variable	Early	Intermediate	Late	*F* (2, 52)
Chronological age	19.0 (1.40)[c]	19.2 (0.65)[c]	23.9 (5.3)[a,b]	4.86[*]
Age of arrival	4.7 (4.31)[b,c]	13.1 (1.49)[a,c]	19.0 (2.78)[a,b]	97.10[**]
Length of residence	13.8 (4.57)[b,c]	5.7 (2.48)[a,c]	5.0 (7.50)[a]	16.57[**]
Years in an L2 school	11.1 (2.42)[b,c]	5.6 (2.58)[a,c]	1.8 (1.94)[a,b]	71.19[**]
Age understood L2	7.3 (3.04)[b,c]	13.5 (1.88)[a]	15.3 (4.20)[a]	32.20[**]
Age spoke L2	7.3 (2.99)[b,c]	13.4 (1.98)[a,c]	16.5 (4.52)[a,b]	36.36[**]
Age conversed in L2	7.8 (3.48)[b,c]	13.7 (2.02)[a,c]	17.6 (4.61)[a,b]	34.97[**]
Pct daily speaks L2	78.1 (17.22)[c]	61.4 (28.37)	49.4 (26.74)	6.24[*]
Pct daily L2 exposure	80.6 (21.05)[b,c]	56.6 (26.71)[a]	44.7 (26.78)	9.45[**]

L2 = English, Pct = percentage; *n*s: early = 19, intermediate = 19, late = 17. Tukey HSD, alpha = .05: [a] differs significantly from early arrivals; [b] differs significantly from intermediate arrivals; [c] differs significantly from late arrivals. [*] *p* < .01, [**] *p* < .001.

proficiency in English (a minimum score of 650 on the Test of English as a Foreign Language; i.e., high intermediate).

Preamble completion task

Materials The experimental preambles (Appendix) were 60 complex noun phrases with a head noun (subject), a local (nonhead) noun embedded in a prepositional phrase, and an adjective or past participle at the end. The AI, IA, and II sequences, exemplified in (12)–(14), were each represented by 20 preambles:

(12) *AI:* The officer(s) from the station(s) . . . suspicious.
(13) *IA:* The letter(s) from the lawyer(s) . . . lost.
(14) *II:* The bridge(s) from the island(s) . . . repaired.

Four different versions of each preamble were created by varying the grammatical number of the head and local nouns (SP, SS, PP, and PS). Each version was assigned to one of four lists, and only one version appeared in each list. This design allowed within-subjects comparisons of animacy and grammatical number. All of the preambles could felicitously be completed with a form of *be*, in which case SP/SS would be expected to elicit *is* or *was*, whereas PP/PS should elicit *are* or *were*. Copular *seem, become, appear,* and so forth, were also expected, but they were much less common than *be* in pilot work. Another 40 items were fillers in the form Determiner + Adjective + Noun . . . Adjective (e.g., *The chubby baby . . . unhappy*). The fillers were balanced for noun animacy and number.

Norming Plausibility, noun frequency, and distributivity were controlled in order to ensure that the experimental items in each animacy condition did not vary with respect to other variables that might influence agreement accuracy. Plausibility ratings were obtained from 20 undergraduate students who rated paired sentences ending in the predicate of the preamble; one sentence started with the head noun of the preamble, and the other started with the local noun:

(15) The bridge was repaired.
(16) The island was repaired.

Plausibility was rated on a 7-point scale, where 7 indicated a *completely acceptable and comprehensible sentence*. Plausibility weights were computed for each dyad by subtracting the mean rating for the sentence with the local-noun subject from the mean for the head-noun version. The weights in the three conditions were comparable (*M*s: AI = 4.02, IA = 3.70, II = 3.95), $F_2(2, 56) < 1$. (Positive weights indicate that the head noun was a more plausible subject than the local noun.)

Noun frequency The SUBTLEX$_{US}$ corpus (Brysbaert & New, 2009) provided word frequency counts for the singular and plural versions of the head and local nouns in each animacy combination (see Table 2 for the mean $\log_{10}$ frequencies). Frequencies were disproportionately higher for singular than for plural nouns [head and local nouns: respective F_2s(1, 57) = 127.69 and 205.20, *p*s < .001], but there were no reliable differences in the mean frequencies of either head or local nouns in each animacy condition [F_2s(2, 57) < 1].

Distributivity Distributive preambles are SP preambles in which the grammatically singular head noun can be construed as denoting more than one instance of the referent, as in (17):

(17) The photo on the postcards was colorful.

Distributivity was controlled because it can enhance SP errors (e.g., Vigliocco, Butterworth, & Garrett, 1996). Sixteen undergraduates decided whether the head noun of

Table 2 Mean $\log_{10}$ frequencies (*SD*s in parentheses) for the singular and plural versions of the head and local nouns in each animacy combination

Animacy	Head Nouns		Local Nouns	
	Singular	Plural	Singular	Plural
AI	2.66 (0.73)	2.10 (0.63)	3.44 (0.58)	2.44 (0.59)
IA	3.02 (0.42)	2.24 (0.44)	3.21 (0.87)	2.56 (0.72)
II	2.85 (0.82)	2.24 (0.61)	3.20 (0.63)	2.36 (0.47)

Frequencies were obtained from SUBTLEX$_{US}$ (Brysbaert & New, 2009). A = animate, I = inanimate.

the SP version of each experimental preamble implied a singular or a plural entity. No reliable differences were apparent between the mean conceptual plurality weights for the three animacy conditions (MAX = 2; AI = 0.07, IA = 0.17, II = 0.11), $F_2(2, 57) = 2.22$, $p = .12$.

Preamble completion procedure The preambles were centered in 16-point bold Times New Roman font on 10.1 × 15.2 cm index cards. Participants were instructed to read the preamble out loud and to complete it orally as quickly as possible. The preambles were manually administered in a different random order to each participant and remained in view while the participant responded. Target preambles that did not elicit a response within 5 s were readministered after the remaining items had been presented. The responses were audiotaped for later verification of the experimenter's written record. A shoulder-level screen prevented the participants from observing what the experimenter wrote.

The memory tasks were presented after the preamble task, as follows. The DF task required participants to repeat sequences of four to eight numbers of increasing sequence length. The DB task required repeating sequences of three to seven numbers in backward order. For the alphabet task, participants heard sequences of unrelated one- or two-syllable concrete nouns, starting with three nouns and progressing incrementally to seven. The nouns of each sequence started with different letters and differed phonologically and semantically. The participants then repeated the sequence alphabetically, according to the first letter of the word (e.g., *mask–arrow–pen* should be repeated as *arrow–mask–pen*).

All three tasks were presented in monotone intonation at a 1-s interitem rate. There were two trials for each sequence length, with different numbers or words on each trial. Two points were credited on each trial when all numbers or words were repeated in the specified order; one point was given when there was an error in order but all numbers or words were reported. No credit was given otherwise. The task ended when the participant failed both trials of a particular sequence length. The maximum score was 20 on each task. The DF

preceded the DB task for half of the participants, and vice versa for the rest; the alphabet task was always last.

Results and discussion

Preamble completion task

Scoring Responses to the experimental preambles were coded as (1) correct agreement, when the verb was inflected for number and agreed in number with the head noun (e.g., *The baker of the cakes was excellent*); (2) agreement error, when an inflected verb disagreed in number with the head noun; this category included subsequently corrected errors (e.g., *The baker of the cakes were . . . was excellent*); or (3) other (e.g., uninflected verbs such as *became*, a change of head or local noun number, no verb, or no response).

Table 3 displays the frequencies in each response category by group, head number, and local noun number. Figures 1 and 2 show the breakdowns of responses by animacy. The visual comparison of the data for early arrivals and native speakers indicated fewer agreement errors in natives, but similar patterns for animacy and grammatical number. In contrast, animacy patterns between early and later arrivals differed markedly. Consequently, the results for natives were compared statistically to early arrivals only. The R software (R Development Core Team, 2011, version R2.12.0) was used to conduct binomial mixed-effect logistic regression analyses (lmer) on verb agreement errors in binary form (incorrect vs. correct; all other response types were entered as "NA"; see Baayen, 2008). In all models, subjects and items were entered as crossed random effects (intercept only).

Early arrivals versus native speakers The predictors in the regression were group (native, early AoA); animacy (AI, IA, II); head number (singular, plural); local noun number; all two-way interactions between group, animacy, and head noun; the Head Noun × Local Noun interaction; and the Group × Animacy × Head Noun interaction. Deviant agreement was less common in natives than in early arrivals [*M*s = 3.53 vs. 8.26; log-odds estimate $B = -1.937$ ($SE = 0.71$), $z = -2.738$, $p < .001$], but animacy was not a significant predictor (all contrasts, $p > .05$). Errors were more common for singular than for plural head nouns (log-odds estimate $B = 1.712$, $SE = 0.38$, $z = 4.448$, $p < .001$) and for plural than for singular local nouns (log-odds estimate $B = 0.788$, $SE = 0.25$, $z = 3.151$, $p < .001$). In addition, the difference between mismatch errors (*M*s: SP = 3.45, PS = 1.39) and match errors (*M*s: SS = 0.26, PP = 0.76) was greater for singular than for plural head preambles (reflected by the Head × Local Noun interaction: log-odds estimate $B = -3.807$, $SE = 0.42$, $z = -9.059$, $p < .001$).

Table 3 Total correct agreement, agreement errors, and proportions of agreement errors in natives and L1-Chinese L2ers by AoA group (SDs in parentheses)

Condition	Response Type			Proportion of Errors
	Correct	Errors	Other[a]	
Natives (n = 19)				
SP	224 (1.40)	42 (1.25)	19 (1.09)	.16 (.03)
SS	263 (1.75)	3 (0.30)	19 (1.09)	.00 (.00)
PP	253 (2.83)	13 (0.75)	19 (1.15)	.05 (.02)
PS	252 (4.17)	9 (0.48)	24 (1.30)	.03 (03)
Early AoA (n = 19)				
SP	174 (4.41)	90 (4.34)	21 (1.9)	.34 (.03)
SS	267 (1.75)	7 (1.06)	11 (1.36)	.02 (.01)
PP	251 (2.79)	16 (1.6)	18 (1.11)	.06 (.02)
PS	220 (4.17)	44 (3.8)	21 (1.86)	.17 (.03)
Intermediate AoA (n = 19)				
SP	182 (3.81)	86 (3.57)	17 (1.59)	.32 (.03)
SS	253 (2.59)	13 (1.79)	19 (1.97)	.05 (.01)
PP	231 (3.04)	30 (2.39)	24 (2.02)	.11 (.02)
PS	194 (2.25)	79 (3.32)	12 (1.44)	.29 (.02)
Late AoA (n = 17)				
SP	205 (3.43)	47 (3.01)	3 (0.53)	.18 (.02)
SS	239 (1.91)	13 (1.63)	3 (0.57)	.05 (.01)
PP	201 (2.79)	50 (2.83)	4 (0.60)	.20 (.01)
PS	174 (3.99)	78 (3.82)	3 (0.30)	.31 (.03)

MAX = 285 for all groups, except late = 255; Proportion of Errors = total errors/(total errors + total correct); AoA = age of arrival in L2-dominant country, S = singular, P = plural. [a] Other = uninflected verbs or noun number change, unscoreable.

Overall AoA and animacy effects The results in Table 3 indicate fewer agreement errors in early arrivals than in later arrivals (consistent with critical-period constraints on acquiring L2 grammars). However, Figs. 1 and 2 indicate that the age-related increase in deviant agreement is restricted to IA and II preambles with a plural head noun. The fact that AoA effects are absent (or minimal) in the case of AI preambles is inconsistent with the critical-period hypothesis, since it implies that early and late L2ers do not differ in their potentials to acquire new grammatical features.

Statistical analyses confirmed the visual impressions above. The predictors were AoA (early, intermediate, late); animacy (AI, IA, II); head number; local noun number; all interactions between group, animacy, and head number; and the Head × Local Noun interaction. The model was pruned stepwise by deleting nonsignificant interactions when likelihood ratio tests indicated that the deletion did not degrade the fit of the model. It should be noted that variables with three levels generate complex models, because two levels are compared to the default (which corresponds to the alphabetically earlier level), but these two levels are not compared with each other. In particular, the regression compared intermediate and late AoAs to early AoAs (the intercept or baseline for AoA), and both IA and II to AI (the intercept for animacy). The contrasts between intermediate and late AoAs and between IA and II were obtained by changing the intercept.

No significant AoA differences in number agreement were elicited by AI preambles (log-odds estimate Bs = 0.513 and 0.528, SEs = 0.40 and 0.41, zs = 1.278 and 1.294, ps = .20). Intermediate and late AoA did not differ reliably from each other in the production of IA and II verb number errors (ps >

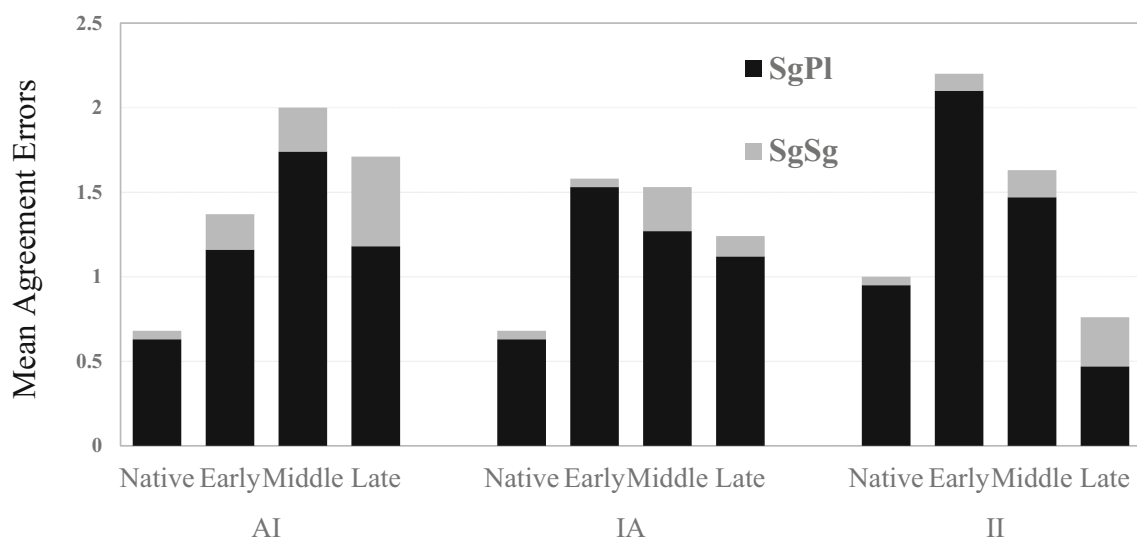## Singular Head Errors by AoA Group in the AI, IA, and II Conditions

Fig. 1 Mean verb number agreement errors by age-of-arrival (AoA) group and animacy (A = animate, I = inanimate) for singular–plural (SgPl) preambles (dark bars) and singular–singular (SgSg) preambles (light bars).
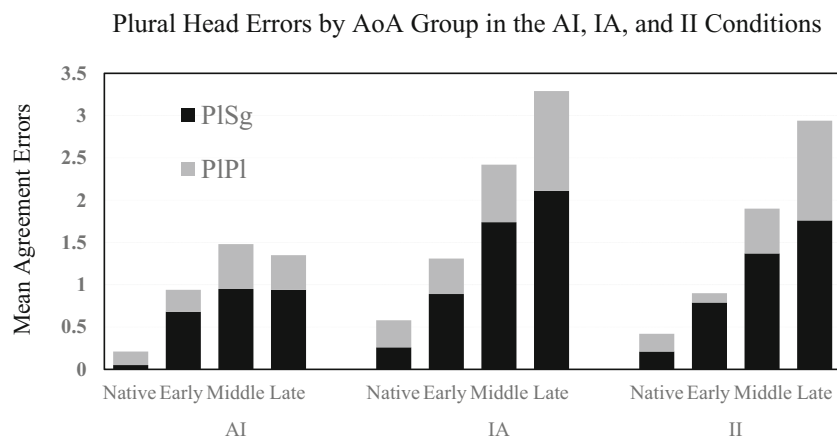
Plural Head Errors by AoA Group in the AI, IA, and II Conditions

Fig. 2 Mean verb number agreement errors by age-of-arrival (AoA) group and animacy (A = animate, I = inanimate) for plural–singular (PlSg) preambles (dark bars) and plural–plural (PlPl) preambles (light bars).

.19), but both groups produced significantly more IA and II errors than early arrivals [log-odds estimate Bs = 0.904 and 1.026, respectively, for intermediate vs. early contrasts for both IA and II (SEs = 0.36 and 0.40, respectively), zs = 2.487 and 2.595, ps = .012 and .009; Bs = 1.259 and 1.473 for late vs. early contrasts for IA and II (SEs = 0.36 and 0.39), zs = 3.452 and 3.738, ps < .001]. Animacy did not reliably influence subject–verb agreement in early arrivals (error Ms: AI = 2.38, IA = 2.78, and II = 3.00; SDs = 2.32, 2.89, and 3.21; ps = .15 to .67). For intermediate arrivals, II (M = 3.53, SD = 1.68) did not differ reliably from AI or IA (respective Ms = 3.47 and 3.90, SDs = 2.5 and 2.04, ps = 22 and .16), but AI elicited fewer errors than IA (log-odds estimate B = 0.727, SE = 0.28, z = 2.557, p = .01). In late arrivals, IA and II did not differ reliably (Ms = 4.31 and 3.70, SDs = 2.05 and 2.15, p = .26), and both elicited significantly more errors than AI (M = 3.06, SD = 2.65) [log-odds estimate Bs = 1.066 and 0.787 (SEs = 0.28 and 0.29), zs = 3.752 and 2.72, ps < .001 and .006]. Finally, mismatch preambles (SP/PS) elicited more errors across AoA than did match preambles (SS/PP), reflected by a significant Head × Local Noun interaction: log-odds estimate B = –3.399 (SE = 0.25), z = –13.707, p < .001.

Effect of head number The next set of contrasts confirmed that the age-related changes in the effect of the number of the head noun were restricted to IA and II; AI errors were comparably greater after singular than after plural head nouns across AoA (Ms for singular and plural: early = 1.37 and 0.95; intermediate = 1.95 and 1.47; late = 1.71 and 1.35; ps > .30). However, in the case of IA, the singular head > plural head asymmetry for early arrivals differed significantly from the head number asymmetry for both intermediate and late arrivals (Ms: early = 1.58 and 1.31 vs. intermediate = 1.48 and 2.42, late = 1.26 and 3.24; log-odds estimate Bs = .877 and 1.753 (SEs = 0.43 and 0.45), zs = 2.052 and 3.901, ps = .04 and <.001). II displayed the same AoA differences in head

number asymmetry as IA (Ms for singular and plural: early = 2.15 and 0.89, intermediate = 1.63 and 1.95, late = 0.78 and 2.90) [log-odds estimate Bs = 1.900 and 2.836 (SEs = 0.84 and 0.83), zs = 2.25 and 3.48, ps = .02 and <.001]. Intermediate AoAs, in turn, differed significantly from late AoAs for both IA and II [Bs = 0.876 and 1.437 (SEs = 0.43 and 0.44), zs = 2.055 and 3.227, ps = .04 and .001].

It is plausible that inefficient processing of plural morphology contributed to the age-related increase in singular verbs after plural head nouns and the concomitant reduction in SP errors seen in Fig. 1. Even so, the absence of reliable AoA differences for AI preambles implies that late L2ers are as capable of acquiring new grammatical features as are early arrivals. The animacy effects revealed in Figs. 1 and 2 and confirmed by the statistical analyses are precisely what would be expected if processing II and IA preambles consumes more resources than is the case for AI, and consequently leaves fewer resources for computing verb number agreement (and/or efficient processing of plural inflections). The next section provides evidence supporting the resource limitation hypothesis, in the form of results from the memory tasks.

Memory tasks

Table 4 contains means for the DF, DB, and alphabet tasks. One-way between-subjects analyses of variance (ANOVAs) indicated no AoA differences for the DB task [$F_1(3, 71) < 1$], but significant group differences for the DF and alphabet tasks [respective F_1s(3, 71) = 9.24 and 3.11, MSEs = 9.33 and 10.58, ps < .001 and .032]. Natives and early arrivals did not differ in forward number recall (p = .24) or alphabetically reordered word recall (p = .99), but both groups surpassed late arrivals on these tasks (for natives, respective ps < .001 and .03; for early arrivals, ps = .007 and .05); intermediates reliably exceeded late arrivals in forward

Table 4 Mean recall scores (*SD*s in parentheses) on digits forward, digits backward, and alphabet tasks in native and L2 English speakers as a function of AoA group

Group	Digits Forward	Digits Backward	Alphabet
Native	16.5 (3.80)	13.7 (3.40)	8.8 (3.85)
Early	14.8 (2.41)	13.9 (4.31)	8.9 (3.06)
Intermediate	13.6 (2.61)	14.2 (3.00)	7.2 (3.37)
Late	11.4 (3.43)	13.8 (4.02)	6.1 (2.13)

See the text for the scoring and results of the statistical tests.

number recall ($p = .02$). The results for the alphabet task, in particular, indicate deficient resources in late arrivals for processing and manipulating L2 verbal materials, relative to early arrivals and native speakers.

Table 5 displays Pearson correlations between the total numbers of agreement errors elicited by IA and II plural head preambles (PS + PP), singular head preambles (SP + SS), the memory tasks, and the demographic variables. Plural head errors were initially examined separately for PS and PP sequences, but subsequently were combined because they correlated highly with each other and generated nearly identical intercorrelation patterns. AI errors were not included because the nonsignificant AoA differences for AI could mask or reduce a possible influence of AoA on agreement errors in the hierarchical regression analyses (see below). Only the Age Conversed variable was included, because the high intercorrelations between the ages at which participants understood, spoke, or conversed in English ($r = .80$ or above) implied all of the variables measured the same factor.

Singular head errors correlated significantly with Age Conversed in English ($p = .04$) and marginally with Years in English schools ($p = .06$) and AoA

($p = .07$). Plural head errors correlated reliably with all three memory tasks; the greater correlation with DF than with DB or alphabet reordering may reflect the wider range of scores for DF. However, plural head errors also correlated significantly with AoA and all other demographic variables, but the relative contributions of the memory tasks and the demographic variables were indeterminable, due to the intercorrelations between them. This confounding was disentangled through hierarchical (ordinary) linear regressions with (II/IA) plural head errors as the dependent variable. The models focused on determining the relative contributions of the memory tasks (resource limitation) and AoA (critical period) to variance in deviant agreement.

In Model 1 (Table 6), DF was entered first because of its high correlation with plural head errors. Four significant partial correlations emerged after entering DF; LOR ($r = -.370, p = .006$), percent of day speaks English ($r = -.312, p = .023$), AoA ($r = .279, p = .043$), and percent of daily L2 exposure ($r = -.277, p = .046$). LOR, entered at Step 2 because of its high partial correlation, predicted an additional 11 % of variance and left DB and percent speaks English as equal candidates for Step 3 (both partial rs $= -.249, p = .068$). DB added 4 % toward explaining variance for IA/II agreement errors; when percent speaks English was added at Step 3 instead of DB, it also accounted for an additional 4 %.

In Model 2, DB and DF were both potential predictors, after LOR was entered at Step 1 (DB, $r = -.372, p = .006$; DF, $r = -.369, p = .007$); partial correlations for current L2 exposure and percent of time speaking English were also reliable at Step 2 (both partial rs $= -.30, p = .03$) but were lower than those for DB/DF. Consequently, DB was entered at Step 2 and accounted for an independent 11 % of

Table 5 Correlation matrix for IA and II verb number errors, the memory tasks, and demographic variables for the L1-Chinese L2ers

Variable	1	2	3	4	5	6	7	8	9	10
1. PlHd	–									
2. SgHd	−.21	–								
3. Digit-F	−.46**	.04	–							
4. Digit-B	−.37**	−.11	.43**	–						
5. Alphabet	−.24*	.11	.54**	.33*	–					
6. AoA	−.43**	−.20	−.48**	−.09	−.31*	–				
7. LOR	−.46**	.05	.34*	.09	.17	−.57**	–			
8. AgeL2Conv	.32*	−.24*	−.37**	−.12	−.34**	.86**	−.48**	–		
9. %SpkL2	−.38**	.15	.25*	.14	.14	−.49**	.30*	−.46*	–	
10. YrsL2Schl	−.37**	.21	.43**	−.04	.35**	−.87**	.70**	−.80**	.50**	–
11. %L2Exp	−.38**	−.11	.35**	.21	.32*	−.50**	.38	.47	.71	.46**

PlHd = PS + PP errors; SgHd = SP + SS errors; AoA = age of arrival in English-dominant country; LOR = length of residence in English-dominant country; AgeL2Conv = age first conversed in English; %SpkL2 = percent of day speaks English versus L1; YrsL2Schl = number of years in English-language school; %L2Exp = percent of day exposed to English versus L1. $^*p < .05$, $^{**}p < .01$.

Table 6 Summary of three different hierarchical regression models examining agreement errors elicited from L1 Chinese by plural head preambles in the IA and II conditions

Step and predictor	R	R^2	ΔR^2	ΔF	df	p
Model 1						
Digit-F	.46	.21	.213	14.06	1, 53	<.001
LOR	.57	.32	.108	8.10	1, 52	.006
Digit-B	.60	.36	.044	3.48	1, 51	.068
Model 2						
LOR	.46	.21	.214	14.12	1, 53	<.001
Digit-B	.57	.32	.109	8.17	1, 52	.006
%SpkL2	.61	.37	.050	3.94	1, 50	.053
Model 3						
AoA	.43	.19	.189	12.13	1, 53	.001
Digit-B	.54	.30	.108	7.85	1, 52	.007
LOR	.60	.36	.061	4.84	1, 50	.032

LOR = cumulative length of residence in English-dominant country; %SpkL2 = percent of day speaks English rather than L1; AoA = age of arrival in English-dominant country.

variance in IA/II plural head agreement errors. Percent speaking English was entered next because the partial correlation for this variable was reliable at Step 3 (partial $r = -.27$, $p = .053$), and it added 5 %.

Model 3 confirmed that DB, DF, and LOR were potential predictors after controlling for AoA at Step 1 (respective partial correlations: DB, $r = -.365$, $p = .007$; DF, $r = -.324$, $p = .018$; LOR, $r = -.295$, $p = .032$). DB added significant variance when entered at Step 2 (almost 11 %); entering LOR at Step 3 (partial $r = -.297$, $p = .032$) explained an additional 6 % of variance in plural head errors, and left no further candidates.

In sum, LOR and DB/DF are reliable predictors of variance in plural head IA/IA agreement errors when AoA is controlled, whereas AoA is not a reliable predictor when DF and LOR are partialed out. These results support the resource limitation account of errors, rather than the critical-period explanation.

General discussion

Elicitation of subject–verb errors was used to test whether the acquisition of subject–verb number agreement by L1 Chinese speakers is subject to maturational constraints or to resource limitations that disrupt access to grammatical knowledge. Verb agreement was reliably more secure overall in native speakers than in early arrivals, and in early than in late arrivals. Although this overall pattern is consistent with what would be

expected if L2 acquisition is constrained by a critical or sensitive period (Johnson & Newport, 1989), significant AoA differences between L2ers were restricted to IA and II preambles with plural head nouns. The absence of reliable AoA differences for AI preambles indicates that late learners can acquire new grammatical features; in particular, it suggests that the mental representations and procedures for computing subject–verb agreement are available to late learners whose native language does not use verb number concurrence. The difficulty with agreement in the case of II and IA preambles and the absence of such an effect for AI are explicable in terms of insufficient resources for handling agreement under high processing cost, and this view was supported by an inverse correlation between scores on all three memory tasks and II/IA errors, as well as by the results of the hierarchical regressions.

The fact that the verb number concordance deficiency was almost entirely restricted to preambles with a plural head noun suggests that it partly resides in handling plural morphology under conditions of high processing cost, as is elaborated in the next section.

Plurality and processing complexity

Bock and Eberhard (1993) and Eberhard (1997) proposed that the lemma for a singular count noun does not carry a grammatical specification for number, whereas the lemma for a plural count noun is tagged with a grammatical feature for number (see also Corbett, 2000, 2006). If the verb agreement mechanism detects an activated number feature on the head noun phrase in the mental representation of the structure of the sentence, it assumes that the noun is plural and retrieves or generates a plural verb form. If no number feature is detected, the agreement device assumes by default that the noun is singular and retrieves or generates a singular verb form.

Plural forms seem to be more complex than singular forms and to be especially likely to heighten cognitive complexity when combined with other sources of complexity (Franck, Vigliocco, & Nicol, 2002). Kail and Bassano (1997) found longer latencies for detecting agreement deviance in French preambles in PP than in SS conditions and in PS than in SP conditions. Fayol et al. (1994, Exp. 2) reported that performing a cognitive task (counting clicks) while transcribing dictated French language noun–noun–verb sentences was more disruptive to subject–verb agreement when both nouns were plural than when they were singular. Consistent with the assumption that dealing with plural nouns is resource-consuming, click-counting errors were greater when head nouns were plural than when they were singular.

Neurophysiological support for the notion that plurals heighten processing costs comes from Kaan (2002), who manipulated clause length and grammatical number (SP, SS, PP, PS) in grammatical and ungrammatical Dutch sentences. The critical clauses were in subject–object–verb order, as in the following ungrammatical SS example (subject and critical verb in italics):

(18) Hoewel volgens het gerucht *de Keizer* de dissident *zullen* gaan
      Although according to the rumor *the emperor* the dissident *will*-plural go
      verbannen is er veel tegenstand.
      ban is there much opposition.

Event-related potential recordings were obtained while participants read the sentences word-by-word and judged their grammaticality. Ungrammatical sentences contained a verb that disagreed in number with the subject (the Dutch version of the example occurs with the plural *zullen* rather than singular *zal*, and thus conflicts with the singular subject *de Keizer*). The immediately pertinent finding concerns the P600 component, a late positive wave that reflects difficulty with syntactic integration (e.g., Featherstone, Gross, Münte, & Clahsen, 2000) and/or processes of repair or reanalysis following detection of an actual or seeming violation (e.g., Friederici, Hahne, & Mecklinger, 1996). As compared to the onset of the P600 for ungrammatical SP sentences (450–500 ms), onset of the P600 was delayed for ungrammatical PS sentences (500–550 ms), and even more delayed for ungrammatical PP sentences (550–600 ms). According to Kaan, plural nouns burden the sentence-processing mechanism more than singular nouns; the added load can draw some resources away from diagnosis and repair, leading to P600 delays for items with more plural nouns. Kaan further suggested that tracking plural features (on the subject) can be affected by linear distance, including the distance between the subject and verb (consistent with Reichle, Tremblay, & Coughlin, 2013). Although generalization from languages with rich inflectional systems (French and Dutch) to English is not possible, the available results are consistent with the notion that plurals are more complex than singulars.

Resource limitation effects on verification and repair
Psycholinguistic models of verb agreement also assume that verification processes are adversely affected when resources are strained. Fayol and collaborators (e.g., Fayol et al., 1999; Fayol et al., 1994; Hupet et al., 1998) proposed a two-step model that imputes subject–verb agreement conflict to inefficient checking and revision. The first step is an automatic procedure of copying the number feature of the noun immediately preceding the verb onto the verb. When this noun is not the head noun, activation may spread from that noun to the verb instead of from the head noun, generating an attraction error. However, the second step is the pertinent one here. At this stage, a checking and revision process verifies the output of the automatic routine, and when conflict is detected, recomputation occurs (consistent with Kaan, 2002). Hartsuiker and Barkhuysen (2006) held that verification targets the level of syntactic integration; when number conflict is detected, verification processes retrieve the grammatical information and may undo an incorrect solution, provided there are sufficient resources. Both Fayol and colleagues and Hartsuiker and Barkhuysen assumed that the verification processes are resource-consuming and that their efficiency is compromised when processing/working memory resources are limited.

Greater difficulty with agreement in IA preambles had been predicted on the assumption that they involve a revision of processing expectations and more intensive semantic–syntactic analysis than AI preambles (which also holds for the II case). Although the difficulty with plurality was not predicted, more intensive, extensive processing of IA/II than of AI preambles would leave relatively fewer resources for processing and/or maintaining number information. In turn, this would increase the likelihood of erroneously generating or retrieving a singular rather than a plural verb. The fact that the subject and verb were separated by the local noun may also have increased the difficulty of tracking plurality on the head nouns (cf. Kaan, 2002; Reichle et al., 2013).

If residual resources for processing plurality were greater in the case of AI than of IA/II, AI preambles should have elicited *more* SP attraction errors than the IA/II sequences, which is precisely what one sees for late arrivals in Fig. 1 (although the difference was not significant). It is also plausible to assume that IA/II plural head number conflicts were more likely to evade detection by checking procedures than were AI conflicts because fewer resources would be available to these processes in the IA/II case. Native speakers and early arrivals may also have found IA/II more difficult than AI preambles, but their larger reserve of L2 resources than are available to later L2ers would allow them to handle plurality without overtly adverse consequences.

A counterargument to this account of the animacy effect is that plurals were easier to process on animate than on inanimate nouns because Mandarin (the version of Chinese for most of the later arrivals) optionally attaches a suffix (-*men*) on plural animate nouns that conceivably would facilitate selective transfer of the notion of a suffix for plural number. In the case of pronouns, the suffix -*men* does denote more than one (e.g., *wo* = I, *wo-men* = we), but researchers dispute whether it denotes plurality, collectivity, or definiteness when added to common nouns (Cheng & Sybesma, 1999;

Iljic, 1994; Y.-H. A. Li, 1999). However, the transfer account could be tested here, because two of the ten plural animate subject fillers were irregular plural nouns ending in -en (children, men). If late arrivals processed -men (or -en) as a marker for plurality, they should have produced fewer verb number errors to these fillers than to the eight regular animate plurals (e.g., babies). Children elicited only one error from late arrivals, but men accounted for 38 % of all animate plural agreement violations.

Implicit versus explicit grammatical knowledge Although the results indicate that late L2ers have difficulty with plural morphology (consistent with Jiang, 2004), their sensitivity to grammatical plurality in the case of AI preambles is inconsistent with Jiang's contention that even proficient L2ers are insensitive to plural morphology during online tasks. Perhaps the materials in Jiang's study were slanted toward inanimate–inanimate sentences (a condition that elicited difficulty with plural morphology in the present work). If so, his results may reflect insufficient resources for efficient processing of plurals (which would explain why his L2 participants took longer to process plural than singular nouns, as was noted earlier). However, Jiang's distinction between implicit and explicit morphosyntactic knowledge in natives and L2ers needs further discussion here (see Bialystok, 1978, for a discussion of the distinction between implicit, unconscious, automatically activated rule knowledge vs. explicit, conscious knowledge; and Paradis, 2004, and Ullman, 2001, for procedural vs. declarative knowledge).

Jiang (2004, 2007) regarded morphosyntactic knowledge in native speakers as being implicit in nature, automatically activated, and applied unconsciously, whereas knowledge in even highly proficient L2ers is explicit and not an automatic part of their L2 competence (for a discussion of automaticity in L2 acquisition, see Segalowitz & Hulstijn, 2005). He argued that implicit grammatical knowledge is essential for online sensitivity to grammatical deviance; although L2ers may do well in offline tasks such as grammaticality judgment tests by applying their explicit knowledge, they may have difficulty with the same linguistic feature or rule during online processing, because they lack implicit L2 knowledge.

Jiang (2004, 2007) also held that L2 speakers do not acquire automatic competence even after many years of immersion in the L2, and pointed to inflectional errors in the two L1-Chinese L2ers studied by Aaronson and Ferres (1987), one of whom was born in the United States, and the other of whom arrived at age 5. Both were highly proficient in English and yet made errors in their use of the past tense and plural. The

greater frequency of agreement errors in the present sample of early arrivals, relative to natives, is superficially consistent with this position, but their large number of singular–plural attraction errors clearly demonstrates sensitivity to plurality during online processing (and, in principle, indicates implicit knowledge).

The large number of errors in early arrivals was unexpected, as they had acquired English informally, through the aural medium, rather than through explicit instruction in the grammar of English, as in late arrivals (and some intermediate arrivals). However, Chinese was their home language during early childhood, and on the average, they still spoke Chinese approximately 20 % of the day. But in fluent bilinguals who use both languages regularly, both languages are active and available, even when only one language is being used. As a group, bilingual adults have lower scores on verbal fluency tasks than do monolinguals, a finding that has been attributed to cross-language interference (e.g., Gollan, Montoya, Fennema-Notestime, & Morris, 2005). If so, the greater frequency of erroneous agreement in early arrivals than in natives could reflect interference from (or competition with) Chinese, and/or a consequent expenditure of resources on inhibiting the L1. The fact that greater current use of English than Chinese contributed to variation in plural head IA/II errors after controlling for LOR and DB further suggests that the more one uses the L2, the less the interference from the L1.

In conclusion, the results of this work indicate that late L2 learners can acquire new grammatical features but have difficulty applying their knowledge under conditions of heavy processing load. In addition, the importance of LOR and of using the L2 seen in the regressions echoes N. C. Ellis (1991, 1998), who stressed the contribution of informal L2 exposure. He regarded linguistic knowledge as derived from a huge collection of memories of previously encountered utterances that allow for the emergence of linguistic regularities as central tendencies in the databases of related memories. According to Ellis, L2 grammars emerge in the form of usage-based, piecemeal learning of grammatical forms and of frequency-based abstractions. Formal instruction (explicit knowledge) is indispensable, but it can impede acquisition when it distorts patterns that occur in naturalistic exposure. Nascent constructions need to be continuously tuned through immersion in the language so that they can become ingrained in the learner's competence and, over time, automatically available (see also Paradis, 2004; Ullman, 2001). The results here are consistent with this perspective; with increased exposure to the L2, fewer resources are needed for processing verbal material, thereby allowing larger chunks of input to be held in memory and encoded, which in turn enhances the likelihood of analyzing out L2 grammatical features.

Appendix

Table 7 Preambles in each animacy condition, in the singular–singular version

Animate–Inanimate	Inanimate–Animate
The officer from the station . . . (suspicious).	The warning from the teacher . . . (alarming).
The boxer from the fight . . . (courageous).	The memo from the accountant . . . (unclear).
The soldier from the battle . . . (brave).	The cheque from the stockbroker . . . (void).
The skater from the contest . . . (graceful).	The message from the secretary . . . (misplaced).
The actor from the play . . . (injured).	The letter from the lawyer . . . (lengthy).
The leader for the march . . . (ambitious).	The penalty for the student . . . (harsh).
The director for the company . . . (angry).	The package for the worker . . . (torn).
The drummer for the parade . . . (tired).	The medicine for the baby . . . (orange).
The professor for the course . . . (handsome).	The pizza for the boy . . . (smelly).
The chef for the dinner . . . (pleased).	The magazine for the model . . . (mailed).
The sprinter for the race . . . (fast).	The trophy for the singer . . . (shiny).
The architect for the building . . . (creative).	The speech for the politician . . . (translated).
The speaker for the rally . . . (boring).	The article for the editor . . . (brief).
The composer for the song . . . (young).	The desk for the judge . . . (clean).
The baker of the cake . . . (fat).	The video of the dancer . . . (cancelled).
The maker of the movie . . . (insane).	The sketch of the girl . . . (skillful).
The owner of the car . . . (rich).	The cartoon of the writer . . . (smeared).
The painter of the mural . . . (talented).	The painting of the President . . . (stolen).
The voter at the election . . . (Canadian).	The repair by the mechanic . . . (costly).
The guard beside the tower . . . (violent).	The bus beside the tourist . . . (dented).

Table 7 (continued)

Animate–Inanimate	Inanimate–Animate

Inanimate–Inanimate

The train from the city . . . (fast).

The wire from the printer . . . (twisted).

The bridge from the island . . . (repaired).

The poster from the museum . . . (framed).

The smell from the truck . . . (horrible).

The design for the airplane . . . (accepted).

The boat for the ride . . . (colourful).

The slide for the lecture . . . (lost).

The contract for the novel . . . (signed).

The switch for the light . . . (sticky).

The piano for the wedding . . . (tuned).

The notebook for the meeting . . . (new).

The hanger for the towel . . . (plastic).

The program for the computer . . . (flawed).

The coupon for the prize . . . (cut).

The photo of the statue . . . (glossy).

The critique of the chapter . . . (fair).

The report of the fire . . . (incorrect).

The map of the lake . . . (outdated).

The rosebush beside the wall . . . (pretty).

References

Aaronson, D., & Ferres, S. (1987). The impact of language differences on language processing: An example from Chinese–English bilingualism. In P. Homil, M. Palij, & D. Aaronson (Eds.), *Childhood bilingualism: Aspects of linguistic, cognitive, and social development* (pp. 75–119). Hillsdale, NJ: Erlbaum.

Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge, UK: Cambridge University Press.

Barker, J., Nicol, J., & Garrett, M. (2001). Semantic factors in the production of number agreement. *Journal of Psycholinguistic Research, 30,* 91–114.

Bialystok, E. (1978). A theoretical model of second language learning. *Language Learning, 28,* 69–84.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language, 44,* 235–249. doi:10.1006/jmla.2000.2750

Bock, J. K. (1986). Syntactic persistence in language production. *Cognitive Psychology, 18,* 355–387. doi:10.1016/0010-0285(86)90004-6

Bock, K., & Eberhard, K. M. (1993). Meaning, sound and syntax in English number agreement. *Language & Cognitive Processes, 8,* 57–99. doi:10.1080/01690969308406949

Bock, J. K., & Miller, C. E. (1991). Broken agreement. *Cognitive Psychology, 45,* 45–93.

Bonin, P., Gelin, M., & Bugaiska, A. (2014). Animates are better remembered than inanimates: Further evidence from word and picture stimuli. *Memory & Cognition, 42,* 370–382. doi:10.3758/s13421-013-0368-8

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods, 41,* 977–990. doi:10.3758/BRM.41.4.977

Cheng, L. L.-S., & Sybesma, R. (1999). Bare and not so bare nouns and the structure of NP. *Linguistic Inquiry, 30,* 509–542.

Corbett, G. G. (2000). *Number.* Cambridge, UK: Cambridge University Press.

Corbett, G. G. (2006). *Agreement.* Cambridge, UK: Cambridge University Press.

Craik, F. I. M. (1986). A functional account of age differences in memory. In F. Klix & H. Hagendoff (Eds.), *Human memory and cognitive abilities* (pp. 409–422). Amsterdam, The Netherlands: Elsevier.

Dewart, M. H. (1979). The role of animate and inanimate nouns in determining sentence voice. *British Journal of Psychology, 30,* 495–501.

Eberhard, K. M. (1997). The marked effect of number on subject–verb agreement. *Journal of Memory and Language, 36,* 147–164. doi:10.1006/jmla.1996.2484

Ellis, R. (1991). Grammaticality judgments and second language acquisition. *Studies in Second Language Acquisition, 13,* 499–533.

Ellis, N. C. (1998). Emergentism, connectionism, and language learning. *Language Learning, 48,* 631–664.

Fayol, M., Hupet, M., & Largy, P. (1999). The acquisition of subject–verb agreement in written French: From novices to experts' errors. *Reading and Writing, 11,* 153–174.

Fayol, M., Largy, P., & Lemaire, P. (1994). Cognitive overload and orthographic errors: When cognitive overload enhances subject–verb agreement errors. A study in French written language. *Quarterly Journal of Experimental Psychology, 47A,* 437–464. doi:10.1080/14640749408401119

Featherstone, S., Gross, M., Münte, T. F., & Clahsen, H. (2000). Brain potentials in the processing of complex sentences: An ERP study of control and raising constructions. *Journal of Psycholinguistic Research, 29,* 141–145.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second language acquisition. *Journal of Memory and Language, 41,* 78–104. doi:10.1006/jmla.1999.2638

Franck, J., Vigliocco, G., & Nicol, J. (2002). Subject–verb agreement errors in French and English: The role of syntactic hierarchy. *Language & Cognitive Processes, 17,* 371–404. doi:10.1080/01690960143000254

Friederici, A. D., Hahne, A., & Mecklinger, A. (1996). Temporal structure of syntactic parsing: Early and late event-related brain potential effects. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22,* 1219–1248. doi:10.1037/0278-7393.22.5.1219

Gerton, B. K., Brown, T. T., Meyer-Lindenberg, A., Kohn, P., Holt, J. L., Olsen, R. K., & Berman, K. F. (2004). Shared and distinctive neurophysiological components of the digits forward and backward tasks as revealed by functional neuroimaging. *Neuropsychologia, 42,* 1781–1787. doi:10.1016/j.neuropsychologia.2004.04.023

Goad, H., & White, L. (2008). Prosodic structure and the representation of L2 functional morphology: A nativist approach. *Lingua, 118,* 577–594.

Goad, H., White, L., & Steele, J. (2003). Missing inflection in L2 acquisition: Defective syntax or L1-constrained prosodic representations? *Canadian Journal of Linguistics, 48,* 243–263.

Gollan, T. H., Montoya, R. I., Fennema-Notestine, C., & Morris, S. K. (2005). Bilingualism affects picture naming but not picture classification. *Memory & Cognition, 33,* 1220–1234. doi:10.3758/BF03193224

Granena, G., & Long, M. H. (2013). Age of onset, length of residence, language aptitude, and ultimate L2 attainment in three linguistic domains. *Second Language Research, 29,* 311–343.

Hartsuiker, R. J., & Barkhuysen, P. N. (2006). Language production and working memory: The case of subject–verb agreement. *Language & Cognitive Processes, 21,* 181–204. doi:10.1080/01690960400002117

Hawkins, R. (2000). Persistent selective fossilization in second language acquisition and the optimal design of the language faculty. *Essex Research Reports in Linguistics, 34,* 75–90.

Hawkins, R. (2001, May). *Linguistic explorations of fossilization in second language acquisition and the critical period hypothesis.* Paper presented at the International Conference on Greek Linguistics, University of Thessaloniki, Greece.

Hawkins, R., & Chan, C. Y.-H. (1997). The partial availability of Universal Grammar in second language acquisition: The "failed functional features hypothesis.". *Second Language Research, 13,* 187–226.

Hawkins, R., & Liszka, S. (2003). Locating the source of defective past tense marking in advanced L2 English speakers. In R. van Hout, H. Aafke, F. Kuiken, & R. Towell (Eds.), *The interface between syntax and lexicon in second language acquisition* (pp. 21–44). Amsterdam, The Netherlands: John Benjamins.

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

Hupet, M., Fayol, M., & Schelstraete, M. A. (1998). Effects of semantic variables on the subject–verb agreement processes in writing. *British Journal of Psychology, 89,* 59–75.

Iljic, R. (1994). Quantification in Mandarin Chinese: Two markers of plurality. *Linguistics, 32,* 91–116.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied PsychoLinguistics, 25,* 603–634.

Jiang, N. (2007). Selective integration of linguistic knowledge in adult second language learning. *Language Learning, 57,* 1–33.

Johnson, J. S., & Newport, E. L. (1989). Critical period effects on second language learning: The influence of maturational state in the acquisition of English as a second language. *Cognitive Psychology, 21,* 60–99.

Kaan, E. (2002). Investigating the effects of distance and number interference in processing subject–verb dependencies: An ERP study. *Journal of Psycholinguistic Research, 31,* 165–193.

Kail, M., & Bassano, D. (1997). Verb agreement in French: A study of on-line grammaticality judgments. *Language and Speech, 40,* 25–46.

Lardière, D. (1998a). Case and tense in the "fossilized" steady state. *Second Language Research, 14,* 1–26.

Lardière, D. (1998b). Dissociating syntax from morphology in a divergent L2 end-state grammar. *Second Language Research, 14,* 359–375.

Lardière, D. (2003). Second language knowledge of [+Past] and [+Finite]. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *Proceedings of the 6th generative approaches to second language acquisition conference (GASLA 2002)* (pp. 176–189). Somerville, MA: Cascadilla Press.

Lardière, D. (2006). Attainment and acquirability in second language acquisition. *Second Language Research, 22,* 239–242.

Lempert, H. (1989). Animacy constraints on preschool children's acquisition of passives. *Child Development, 60,* 237–245.

Lempert, H. (1990). Acquisition of passives: The role of animacy, salience, and lexical accessibility. *Journal of Child Language, 17,* 677–696.

Lenneberg, E. H. (1967). *Biological foundations of language.* New York, NY: Wiley.

Li, Y.-H. A. (1999). Plurality in a classifier language. *Journal of East Asian Linguistics, 8,* 229–242.

Li, C. N., & Thompson, S. A. (1989). *Mandarin Chinese: A functional reference*. Oakland, CA: University of California Press.

Liu, H., Bates, E., & Li, P. (1992). Sentence interpretation in bilingual speakers of English and Chinese. *Applied PsychoLinguistics, 13,* 451–484.

Matthews, S., & Yip, V. (2011). *Cantonese: A comprehensive grammar*. New York, NY: Routledge.

McDonald, J. L. (2006). Beyond the critical period: Processing-based explanations for poor grammaticality judgments by late second language learners. *Journal of Memory and Language, 55,* 381–401. doi:10.1016/j.jml.2006.06.006

Miyake, A., & Friedman, N. P. (1998). Individual differences in second language proficiency: Working memory as language aptitude. In A. E. Healey & L. J. Bourne Jr. (Eds.), *Foreign language learning: Psycholinguistic studies on training and retention* (pp. 339–364). Mahwah, NJ: Erlbaum.

Paradis, M. (2004). *A neurolinguistic theory of bilingualism*. Amsterdam, The Netherlands: John Benjamins.

Pearlmutter, N., Garnsey, S. M., & Bock, K. E. (1999). Agreement processes in sentence comprehension. *Journal of Memory and Language, 41,* 427–456.

Penfield, W., & Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.

Prévost, P., & White, L. (2000). Missing surface inflection or impairment in second language acquisition? Evidence from tense and agreement. *Second Language Research, 16,* 103–133.

R Development Core Team. (2011). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. **Retrieved from www.R-project.org**

Rakison, D. H., & Poulin-Dubois, D. (2001). Developmental origin of the animate–inanimate distinction. *Psychological Bulletin, 127,* 209–228. doi:10.1037/0033-2909.127.2.209

Reichle, R. V., Tremblay, A., & Coughlin, C. E. (2013). Working-memory capacity effects in the processing of non-adjacent subject–verb agreement: An event-related brain potentials study. In *Selected proceedings of the 2011 Second Language Research Forum* (pp. 54–69). Somerville, MA: Cascadilla Proceedings Project.

Segalowitz, N., & Hulstijn, J. (2005). Automaticity in bilingualism and second language learning. In J. F. Kroll & A. M. B. de Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 371–388). New York, NY: Oxford University Press.

Solomon, E. S., & Pearlmutter, N. J. (2004). Semantic integration and syntactic planning in language production. *Cognitive Psychology, 49,* 1–46. doi:10.1016/j.cogpsych.2003.10.001

Su, I. R. (2001a). Context effects on sentence processing: A study based on the competition model. *Applied PsychoLinguistics, 22,* 167–189.

Su, I. R. (2001b). Transfer of sentence processing strategies: A comparison of L2 learners of Chinese and English. *Applied PsychoLinguistics, 22,* 83–112.

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

Vigliocco, G., Butterworth, B., & Garrett, M. F. (1996). Subject–verb agreement in Spanish and English: Differences in the role of conceptual constraints. *Cognition, 61,* 261–298.

Wang, L., Schlesewsky, M., Bickel, B., & Bornkessel-Schlesewsky, I. (2009). Exploring the nature of the "subject"-preference: Evidence from the online comprehension of simple sentences in Mandarin Chinese. *Language & Cognitive Processes, 24,* 1180–1226. doi:10.1080/01690960802159937

White, L. (2010). Second language acquisition at the interfaces. *Lingua, 121,* 577–590.

Xuefang, L. (2010). Influence of Chinese topic-prominent structures on the acquisition of four types of English subject. *Chinese Journal of Applied Linguistics, 33,* 80–94.