Studies in Second Language Acquisition **40** (2018), 907–921
doi:10.1017/S0272263118000025

Research Report

THE PRODUCTION OF SUBJECT-VERB AGREEMENT AMONG SWEDISH AND CHINESE SECOND LANGUAGE SPEAKERS OF ENGLISH

Carrie N. Jackson*

The Pennsylvania State University

Elizabeth Mormer

The Pennsylvania State University

Laurel Brehm

Max Planck Institute for Psycholinguistics, Nijmegen, NL

Abstract

This study uses a sentence completion task with Swedish and Chinese L2 English speakers to investigate how L1 morphosyntax and L2 proficiency influence L2 English subject-verb agreement production. Chinese has limited nominal and verbal number morphology, while Swedish has robust noun phrase (NP) morphology but does not number-mark verbs. Results showed that like L1 English speakers, both L2 groups used grammatical and conceptual number to produce subject-verb agreement. However, only L1 Chinese speakers—and less-proficient speakers in both L2 groups—were similarly influenced by grammatical and conceptual number when producing the subject NP. These findings demonstrate how L2 proficiency, perhaps combined with cross-linguistic differences, influence L2 production and underscore that encoding of noun and verb number are not independent.

Agreement is a core component in many languages. While traditional grammars often treat subject-verb number agreement as a straightforward process—verbs are marked

We thank Marianne Gullberg at Lunds Universitet in Sweden for help with collection of the L1 Swedish speaker data. We also thank Jack DiMidio, Marta Millar, Alexa Rossi, and Ted Smith for help with the L1 Chinese and L1 English speaker data collection and transcription, and two reviewers for their valuable comments. This research was funded in part by the National Science Foundation under grant OISE-0968369 (PI: J. F. Kroll; co-PIs: P. E. Dussias and J. G. van Hell).

*Correspondence concerning this article should be addressed to Carrie N. Jackson, 442 Burrowes Building, Department of Germanic and Slavic Languages and Literatures, Pennsylvania State University, University Park, PA 16802. E-mail: cnj1@psu.edu

plural for plural subjects and singular for singular subjects—speakers consider more than the grammatical number of the subject noun phrase (NP) when computing agreement (e.g., Vigliocco & Hartsuiker, 2002). Additional factors include semantic information (i.e., conceptual number), morphophonological ambiguity, and the distribution of singular versus plural agreement for an NP (e.g., collective vs. noncollective nouns) in a language (e.g., Haskell, Thornton, & MacDonald, 2010; Lorimor, Jackson, & Foote, 2015).

Because agreement involves the rapid integration of multiple information sources, investigating how L2 speakers produce agreement provides critical insight into the cognitive mechanisms underlying L2 production (e.g., Foote, 2010; Hatzidaki, Branigan, & Pickering, 2011; Hoshino, Dussias, & Kroll, 2010; Hoshino, Kroll, & Dussias, 2012; Nicol & Greth, 2003; Wei, Chen, Liang, & Dunlap, 2015). Key questions include whether L2 speakers rely on similar weightings of grammatical versus conceptual information in both languages and whether the presence of subject-verb agreement in the L1 influences L2 agreement. Testing these provides a means of addressing core issues in L2 research regarding whether L2 and L1 speakers rely on similar mechanisms and the degree of L2 influence from L1 morphosyntactic features (e.g., Clahsen & Felser, 2006; Cunnings, 2017; MacWhinney, 2012).

Existing evidence for the role of L1 number morphology in L2 production comes from Japanese and Chinese L2 English speakers, who have difficulty producing and comprehending English subject-verb agreement. This is usually attributed to the limited (and optional) use of noun number morphology and the absence of verb number morphology in their L1 (Chen, Shu, Liu, Zhao, & Li, 2007; Hoshino et al., 2012; Jiang, 2004; Wei et al., 2015). In contrast to Chinese, Modern Swedish exhibits robust NP morphology, marking nouns for number and grammatical gender, but does not number-mark verbs: The verb form is identical for singular and plural NPs. Contrasting L2 English production from L1 speakers of these two languages—and investigating whether L2 proficiency impacts agreement independent of a speaker's L1—provides unique insight on whether L1 number morphology and L2 proficiency influence the use of grammatical versus conceptual number information in L2 production.

GRAMMATICAL AND CONCEPTUAL NUMBER IN L1 PRODUCTION

For languages that number-mark nouns and verbs, grammatical number encoding in the subject NP is essential for agreement. Many experimental studies investigating subject-verb agreement in language production use a variant of a sentence completion paradigm (Bock & Miller, 1991), where participants read or hear a sentence preamble (i.e., the subject NP), like (1) or (2), and then repeat the preamble and complete the sentence.

(1) The vase with the rose (Sing.-Sing.)
(2) The vase with the roses (Sing.-Pl.)

Cross-linguistically, speakers are more likely to make subject-verb agreement errors by using a plural verb with a singular subject NP (e.g., *The vase with the roses were fragile*) in sentences like (2), which contains a number mismatch between the grammatical number of the head noun (i.e., *vase*) and the local noun (i.e., *roses*), than with sentence

preambles containing no such mismatch, as in (1)—a phenomenon referred to as "attraction" (see Lorimor, Bock, Zalkind, Sheyman, & Beard, 2008, for review of cross-linguistic findings).

Conceptual number also impacts agreement such that speakers are more likely to produce plural agreement with subject NPs like (4) than (3) (see Lorimor et al., 2008, for review).

(3) The vase with the roses (Sing.-Pl.; Single Referent)
(4) The label on the bottles (Sing.-Pl.; Distributive Referent)

Although both head nouns are grammatically singular, (4) has a distributive (multiple referent) reading of the subject NP, with a separate label on each bottle. In contrast, the vase in (3) most likely refers to a single object containing multiple roses, leading to a nondistributive (single referent) interpretation.

Several dominant production models can successfully account for the interaction between conceptual and grammatical number (e.g., Eberhard, Cutting, & Bock, 2005; Haskell et al., 2010). In the marking and morphing model (Eberhard et al., 2005), speakers retrieve conceptual and grammatical number information about the target referent. Potential conflicts are resolved at the point of syntactic integration, such that both conceptual and grammatical factors are used in L1 production to determine the choice of a singular or plural verb (see Haskell et al., 2010 for an alternate explanation within a constraint-based framework).

Traditionally, L1 production research has focused on the production of agreement marking on the verb—any number-marking errors on subject NPs are attributed to faulty memory of the preamble and considered distinct from processes involved in agreement (e.g., Bock & Cutting, 1992; Gillespie & Pearlmutter, 2011; but see Thornton & MacDonald, 2003). A recent connectionist model (Brehm & Goldrick, 2016) suggests that both types of errors have the same underlying cause: Coactivation of similar representations leads speakers to alter the number marking of the subject NP (producing e.g., *The label<u>s</u> on the bottles*) and to use inflected verbs that match the local noun, resulting in attraction. This is consistent with other L1 and L2 sentence processing models, like the good-enough framework (e.g., Patson & Husband, 2016; see Christianson, 2016, for review) or the noisy channel model (e.g., Gibson, Bergen, & Piantadosi, 2013; Levy, 2008), which suggest that speakers build incomplete or nonveridical representations during comprehension that result in reconstrual of the input. Each of these frameworks underscores the importance of the encoding and retrieval of subject NP number for the comprehension of subject-verb agreement, matching data from L1 production (e.g., Gillespie & Pearlmutter, 2011; Slevc & Martin, 2015; see Lorimor et al., 2015 for review of retrieval phenomena). These difficulties are likely exacerbated among L2 speakers, who may face additional interference from competing L1 morphosyntactic features along with greater difficulties in retrieving critical information from memory during production (see Cunnings, 2017, for parallel discussion in L2 comprehension).

GRAMMATICAL AND CONCEPTUAL NUMBER IN L2 PRODUCTION

Nicol and Greth (2003) reported that late-learning L2 Spanish speakers produced nearly identical rates of agreement errors in English and Spanish, with few agreement errors on

singular-singular subject NPs regardless of distributivity. They were also sensitive to conceptual number, with more agreement errors on distributive than nondistributive singular-plural subject NPs. There was also a significant correlation between individual participants' error rates in L1 English and L2 Spanish, leading Nicol and Greth to conclude that L2 speakers rely on similar mechanisms for computing subject-verb agreement in both languages. Similarly, Foote (2010) found that both grammatical and conceptual number influence subject-verb agreement production among L2 Spanish (English L1) and L2 English (Spanish L1) speakers in the L1 and the L2, regardless of proficiency and age of acquisition (but see Hoshino et al., 2010, for proficiency-related effects in the use of conceptual number in L2 English).

In a study contrasting Japanese and Spanish L2 English speakers, Hoshino, Kroll, and Dussias (2012) suggest that the absence of number marking and agreement in the L1 can limit the influence of conceptual number in L2 English. Both speaker groups were sensitive to grammatical number, and advanced Spanish L2 English speakers were sensitive to conceptual number. However, Japanese L2 English speakers—even when matched in L2 English proficiency with the L1 Spanish speakers—exhibited no distributivity effects in sentence production despite being sensitive to conceptual number in an offline rating task. Hoshino et al. (2012) attribute these differences to the fact that Japanese does not overtly mark nouns or verbs for number. The absence of L1 processing routines for computing agreement heightened the cognitive load associated with subject-verb agreement in Japanese L2 English speakers relative to Spanish L2 English speakers, leading them to overuse the more reliable cue of grammatical number (see also Hartsuiker, Kolk, & Huinck, 1999). Similarly, cross-linguistic differences in grammatical number specification (e.g., *trousers* is plural in English but singular in Greek) can modulate the computation of L2 subject-verb agreement, providing additional evidence of cross-linguistic influence during L2 production (Hatzidaki et al., 2011).

Wei et al. (2015) provide further evidence that the absence of L1 number morphology can lead to difficulties integrating conceptual and grammatical number in L2 production. In a task similar to Hoshino et al. (2010, 2012), conceptual number failed to modulate highly proficient Chinese L2 English speakers' agreement patterns. However, when participants were instead prompted to describe pictures depicting the same preambles, highlighting their conceptual number, L1 Chinese speakers produced more plural agreement with distributive than nondistributive referents (see also Eberhard, 1999; Foote, 2010). This underscores that while Chinese L2 English speakers may not use conceptual number like L1 English speakers, it remains critical to their conceptualization and encoding of number.

PRESENT STUDY

By contrasting Swedish and Chinese L2 English speakers, the present study conceptually replicates research on the impact of grammatical and conceptual number on L2 agreement production, while expanding this research's scope through the inclusion of a novel L1-L2 pairing (Swedish L2 English speakers) and the analysis of both verb- and noun-number marking errors. In so doing, this study provides the opportunity to illuminate the origin of nonnativelike agreement patterns in L2 English by posing the following research questions:

1. Are Chinese and Swedish L2 English speakers similarly influenced by conceptual and grammatical number when producing verb-number marking in L2 English?
2. Are Chinese and Swedish L2 English speakers similarly influenced by conceptual and grammatical number when producing noun-number marking in L2 English?
3. Are verb- or noun-number marking errors modulated by L2 English proficiency?

Based on previous work (Hoshino et al., 2012; Wei et al., 2015), we hypothesize that L1 Chinese speakers should have difficulty using conceptual number in L2 English, leading to few differences in the proportion of agreement errors on the verb for distributive versus single referent subject NPs. If robust L1 NP number morphology is sufficient for speakers to use grammatical and conceptual number when producing L2 subject-verb agreement, Swedish L2 English speakers should produce agreement in a manner similar to L1 English speakers (Eberhard, 1999; Humphreys & Bock, 2005); if L1 verbal morphology is critical, then both L2 groups should show few differences in the proportion of agreement errors for distributive versus single referent NPs. Preamble repetition errors provide additional evidence on the roles of memory (e.g., Cunnings, 2017) and L1 morphosyntax (e.g., Hatzidaki et al., 2011) in L2 production. We predict that noun-number errors may increase for distributive referents due to difficulty with number encoding, especially in the presence of a local plural noun. If L1 noun number morphology supports L2 number encoding, Chinese L2 English speakers should show more noun-number errors in preamble repetition than Swedish L2 English speakers. Finally, we predict that L2 proficiency may modulate the proportion of noun- and verb-number errors, especially for distributive referents (Hoshino et al., 2010), providing further insight into the link between L1 and L2 production.

METHOD

PARTICIPANTS

Twenty-seven Swedish L2 English speakers from the student population at a Swedish university, and 39 Chinese L2 English speakers and 43 L1 English speakers from the student population at an American university participated for payment. Three Swedish speakers, four Chinese speakers, and three L1 English speakers were excluded because they grew up in bilingual households. Six L1 English speakers were excluded because they reported nine months or more of L2 immersion experience and three additional L1 English speakers were excluded due to recording errors resulting in data loss. An additional seven Chinese speakers were excluded because they produced less than 20 fluent experimental items during the oral sentence completion task. The final participant pool included 24 Swedish L2 English speakers (12 female; 12 male), 28 Chinese L2 English speakers (14 female; 14 male), and 26 L1 English speakers (14 female, 12 male).[1] The L2 speakers also completed a 43-point multiple choice proficiency test probing their English grammatical accuracy, with items taken from the Michigan English Language Institute College Entrance Test (MELICET), which is a retired version of the Michigan test for English as a Second Language (used with permission from Blattner, 2007; based on data from the present study, Cronbach's $\alpha = .844$).

See Table 1 for biographical information for all participants. The L1 Chinese speakers were significantly younger than the L1 Swedish speakers ($t(50) = 2.08$, $p = .042$), although the average age in both groups was below 30 years and the overall age range was similar across groups. There was no significant difference in the age at which the L1 Swedish and L1 Chinese speakers began learning English ($t(32.6) = 0.10$, $p = .921$), which was at school between the ages of 5 and 16.[2] However, the L1 Swedish speakers were more proficient than the L1 Chinese speakers on the English proficiency test ($t(50) = 6.21, p < .001$).[3] Finally, the L1 Chinese speakers had spent significantly longer living in an English-speaking country than the L1 Swedish speakers ($t(46) = 2.28$, $p = .027$).[4]

MATERIALS

Data were collected through an oral sentence completion task (Bock & Miller, 1991) using sentence preambles (i.e., subject NPs) based on previous studies (Antón-Méndez & Hartsuiker, 2010; Foote, 2010; Hoshino et al., 2010; see IRIS for full materials at www.iris-database.com). Preambles contained a singular head noun followed by a prepositional phrase, varying according to whether the local noun was singular, as in (1), or plural, as in (2). Twenty items were classified as having a single referent (i.e., conceptually and grammatically singular), as in (3), and 20 were classified as having a distributive referent (i.e., conceptually plural but grammatically singular), as in (4).

To confirm variance in distributivity, 36 L1 English speakers—none of whom participated in the main experiment—were recruited using Mechanical Turk to rate whether items referred to one thing or more than one thing (Foote, 2010). These participants rated the 40 singular-plural experimental items, plus 10 complex NPs containing singular head and local nouns (e.g., *The light in the kitchen*) and 10 complex NPs containing a plural head noun and a singular local noun (e.g., *The tools in the drawer*). Ratings were coded as 1 for a response of "one thing" and 2 for a response of "more than one thing." A *t*-test revealed that the distributivity ratings for single referent items ($M = 1.08$) were significantly lower than those for distributive referent items ($M = 1.23$; $t(38) = 4.81$, $p < .001$).

The 40 experimental items were divided into two lists. Participants saw 10 items per condition and only one version of any item. Items were presented in a pseudo-randomized order along with 60 filler items, consisting of 20 conjoined NPs (e.g., *The chair and the stool*), 30 simple NPs (e.g., *The grey elephants*), and 10 complex NPs with a plural head noun and either a plural or singular local noun (e.g., *The pictures in the album*).

TABLE 1. Biographical information

	Swedish L2 English Speakers		Chinese L2 English Speakers		L1 English Speakers	
	M (SD)	*range*	*M (SD)*	*range*	*M (SD)*	*range*
Current age (years)	25.3 (4.4)	20–38	22.9 (3.9)	18–35	20.5 (1.9)	18–26
Age of first L2 exposure (years)	9.2 (1.3)	7–12	9.1 (2.9)	5–16	NA	
L2 English proficiency (max. 43)	38.0 (4.3)	27–43	29.7 (5.2)	19–38	NA	
Length of residence in L2 country (months)	8.3 (25.6)	0–114	24.5 (23.8)	1–96	NA	

PROCEDURE

Participants completed the experiment individually on a computer using E-prime (Schneider, Eschman, & Zuccolotto, 2002). L1 and L2 participants completed the oral sentence completion task and a self-paced reading task (order counterbalanced across participants) to measure their comprehension of similar sentences, the results of which are not reported here.[5] L1 and L2 participants then completed a language background questionnaire and the L2 English speakers additionally completed the English proficiency test.

For the oral sentence completion task, each trial began with a fixation point (see Figure 1). When the participant pressed a key on the button box, this was replaced with an adjective, displayed for 700 milliseconds. Then the sentence preamble (i.e., subject NP) was displayed for 2,500 milliseconds. Following the preamble, participants heard a tone to signal that they should combine the adjective with the preamble to produce a complete sentence (e.g., *The label on the bottles is/are red*). Once the participant completed the sentence, the experimenter clicked the mouse to initiate the following trial.

Each participant completed seven practice trials. During this time, they received feedback on the correct procedure but no feedback regarding grammatical agreement. During the main portion of the task, no feedback was provided. Responses were recorded with a digital recorder.

SCORING

A total of 3,120 responses (960 from the Swedish L2 English speakers; 1,120 from the Chinese L2 English speakers; 1,040 from the L1 English speakers) were transcribed and coded as correct preambles produced with a singular or plural verb, or as miscellaneous responses. All data were transcribed and coded by L1 English-speaking research assistants and checked by the first author for accuracy. If participants used an adjective that was sensible for the preamble, it was marked as correct even if it did not match the prompt. Miscellaneous responses included instances where participants did not correctly repeat the preamble, hesitated, or repeated part of the preamble; paused or inserted a hesitation marker between the preamble and the verb; responded with a verb that was unmarked for number; or failed to complete the sentence. Miscellaneous responses were excluded from the verb production analyses.[6]

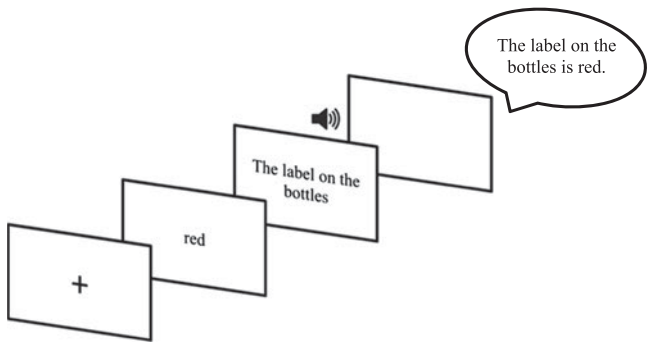

FIGURE 1. Illustration of oral sentence completion trial.

RESULTS

Prior to all analyses, one distributive referent item (*the waste in the river(s)*) was removed because it contained a head noun that would be highly marked in its plural form in the context presented in this experiment. To address RQ1, the verb production data were analyzed using mixed-effect logistic regression models with the lme4 package version 1.1-12 (Bates, Mächler, Bolker, & Walker, 2015) in R version 3.2.5 (R Development Core Team, 2016). Participants rarely produced plural agreement on items with a singular local noun. Thus, only singular-plural items were included in statistical comparisons. Number marking on the verb (singular vs. plural) was the dependent variable. Group (L1 English speaker vs. Swedish L2 English speaker vs. Chinese L2 English speaker) was coded using Helmert contrasts, with the first comparison being between the L1 English speakers and the L2 speaker groups (L1 speaker vs. L2 speaker) and the second comparison being between the L2 speaker groups (L1 Swedish vs. L1 Chinese). As conceptual number is continuous, we operationalized this variable as an item's mean rating from the distributivity rating task, centered at the sample mean. All models used the maximal random effect structure supported by the experimental design (Barr, Levy, Scheepers, & Tily, 2013), including random intercepts for items and participants, decorrelated by-participant random slopes for distributivity and decorrelated by-item random slopes for group.[7]

Table 2 and Figure 2 present descriptive results for the verb production data. As seen in Table 3, there was no significant distributivity by group interaction, so we ran a second model including only main effects. Here, there was a significant effect of distributivity because the proportion of plural responses increased for items rated as more notionally plural (distributive referent items). There was a significant effect of group because the L2 speakers produced more plural responses than the L1 speakers overall, as indicated by the large parameter estimate for the L1 speaker versus L2 speaker comparison within the variable group. Critically, there was no significant difference in the proportion of plural responses between the two L2 speaker groups, as indicated by the small parameter estimate for the Swedish versus Chinese comparison within the variable group.

To address RQ2, the noun preamble production data were analyzed, focusing on miscellaneous responses that were preamble repetition errors with incorrect noun-number marking (e.g., saying *the label<u>s</u> on the bottles*). The L1 English speakers made few miscellaneous errors overall (see Table 3), averaging less than one noun-number marking error per condition (see Figure 3) so their data are not included. As the number of miscellaneous errors for the L1 Swedish and L1 Chinese speakers varied substantially by condition, we compared between-group effect sizes and confidence intervals (CIs) rather than using mixed-effect logistic regressions (Cumming, 2014). Using the BootES package in R (Kirby & Gerlanc, 2013), we calculated the proportion of number-marking and other miscellaneous errors as a function of the total number of errors within each condition for each participant. We then generated bootstrapped 95% CIs and effect sizes using a standardized scale and Hedge's *g* to compare each condition and error type between the L1 Swedish and L1 Chinese speakers.

Figure 3 presents descriptive results for noun-number marking errors and other miscellaneous errors. As seen in Table 4, the only reliable difference between Chinese and Swedish L2 English speakers—indicated by CIs that do not span zero—was in the

TABLE 2. Distribution of responses by experimental condition and group (percentages in parentheses)

	Singular-Singular		Singular-Plural	
	Single Ref.	Distributive Ref.	Single Ref.	Distributive Ref.
Swedish L2 English speakers				
Correct agreement (Sg.)	183 (76.3%)	183 (81.0%)	147 (61.2%)	111 (48.2%)
Agreement error (Pl.)	3 (1.2%)	3 (1.3%)	28 (11.7%)	54 (23.5%)
Unmarked verb	4 (1.7%)	1 (0.4%)	1 (0.4%)	2 (0.9%)
Miscellaneous	50 (20.8%)	39 (17.3%)	64 (26.7%)	63 (27.4%)
(misc. plural, misc. singular)	(16, 1)	(10, 0)	(20, 5)	(13, 12)
Chinese L2 English speakers				
Correct agreement (Sg.)	193 (68.9%)	194 (72.4%)	142 (50.7%)	104 (39.4%)
Agreement error (Pl.)	5 (1.8%)	8 (3.0%)	39 (13.9%)	52 (19.7%)
Unmarked verb	0 (0.0%)	0 (0.0%)	0 (0.0%)	1 (0.4%)
Miscellaneous	82 (29.3%)	66 (24.6%)	99 (35.4%)	107 (40.5%)
(misc. plural, misc. singular)	(8, 0)	(15, 0)	(13, 2)	(4, 9)
L1 English speakers				
Correct agreement (Sg.)	232 (89.2%)	232 (93.9%)	212 (81.5%)	188 (77.1%)
Agreement error (Pl.)	0 (0.0%)	0 (0.0%)	8 (3.1%)	25 (10.2%)
Unmarked verb	7 (2.7%)	5 (2.0%)	3 (1.2%)	4 (1.6%)
Miscellaneous	21 (8.1%)	10 (4.1%)	37 (14.2%)	27 (11.1%)
(misc. plural, misc. singular)	(7, 1)	(1, 0)	(6, 0)	(3, 0)

Note: Under miscellaneous responses, the subset of responses that remained codable for verb-number marking are reported in parentheses.

proportion of noun-number marking errors for distributive referent items containing a plural local noun. This difference stems from the L1 Chinese speakers making more noun-number marking errors in this condition than the L1 Swedish speakers. There were no reliable group differences in noun-number marking errors in the remaining three conditions or in the proportion of other miscellaneous errors in any condition.

To address RQ3 we probed for significant correlations between L2 proficiency, as measured by participants' score on the 43-point English proficiency task, and the number of subject-verb agreement errors and noun-number preamble errors produced. To

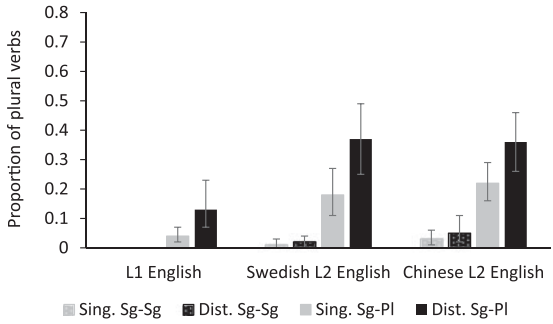

FIGURE 2. Proportion of plural-marked verbs (error bars represent the bootstrapped by-participant 95% CIs).

TABLE 3. Summary of the mixed logit model for singular versus plural verb production

Predictor	Parameter Estimates		Wald's Test		Δ(-2Λ)-test	
Fixed effects	Estimate	Std. error	z-value	Pr (>|z|)	χ^2	*p*
Full model						
(Intercept)	-1.89	0.20	-9.50	<.001		
Distributivity	3.20	1.23	2.62	.009		
Group						
L1 vs. L2	2.41	0.48	5.05	<.001		
Swedish x Chinese	0.19	0.34	0.57	.568		
Distributivity × Group					1.30	.522
Distributivity × Group (L1 vs. L2)	-1.31	2.42	-0.54	.588		
Distributivity × Group (Swedish vs. Chinese)	-1.69	1.72	-0.98	.325		
Main effects only						
(Intercept)	-1.88	0.20	-9.52	<.001		
Distributivity	3.06	1.21	2.54	.011	5.76	.016
Group					27.19	<.001
L1 vs. L2	2.37	0.47	5.02	<.001		
Swedish × Chinese	0.17	0.34	0.50	.614		

increase statistical power, this analysis pooled both L2 groups together. There was a significant correlation between L2 proficiency and the number of noun-number marking errors with distributive referent items containing a plural local noun (r = -.410, n = 52,

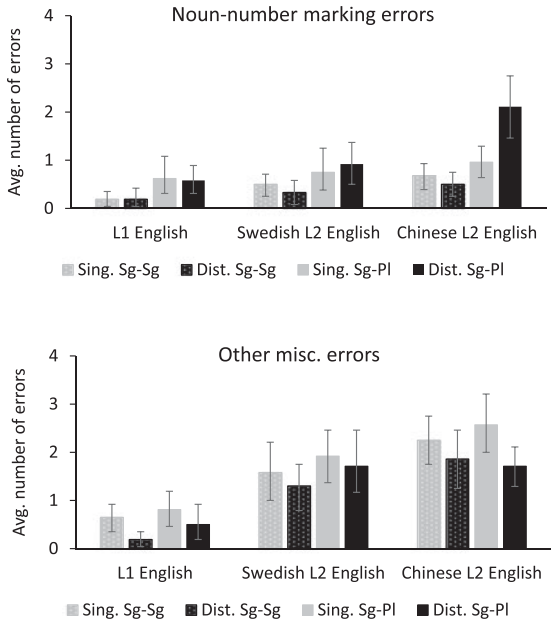

FIGURE 3. Distribution of noun-number marking errors and other miscellaneous errors (error bars represent the bootstrapped by-participant 95% CIs).

$p = .003$). As seen in Figure 4, the number of noun-number marking errors decreased as L2 proficiency increased. No other correlations were statistically significant (all $ps > .12$).[8] Visual inspection of Figure 4 suggests that proficiency differences in both L2 groups contributed to this significant correlation, with less-proficient speakers generally making more errors than more-proficient speakers in both groups. While the L1 Chinese group tended to be more variable, L1 Chinese and L1 Swedish speakers with overlapping L2 proficiency scores made a similar number of errors.

DISCUSSION

In the present study, Chinese and Swedish L2 English speakers exhibited significant attraction effects, producing more subject-verb agreement errors with singular-plural subject NPs than with singular-singular subject NPs. These effects were modulated by conceptual number, with participants producing the highest proportion of errors on singular-plural subject NPs with a distributive referent interpretation (e.g., *The label on the bottles*). An analysis of miscellaneous errors revealed that the L1 Chinese speakers made more noun-number marking errors (e.g., *label* → *label<u>s</u>*) on distributive referent items with a plural local noun than the L1 Swedish speakers, whereas the proportion of other miscellaneous errors was similar for the two L2 speaker groups. Combined across both L2 speaker groups, noun-number marking errors in this condition also significantly correlated with L2 proficiency.

In previous studies, the use of conceptual number in L2 agreement has emerged for L2 speakers whose L1 marks both nouns and verbs for number (Foote, 2010; Hoshino et al., 2010; Nicol & Greth, 2003), and studies that directly compare L2 speakers with different L1s have found that such conceptual number effects are largely limited to those L2 speakers whose L1 marks nouns and verbs for number (Hoshino et al., 2012; Wei et al., 2015). In contrast, both L2 groups in the present study successfully integrated conceptual and grammatical number for subject-verb agreement, despite the varying morpho-syntactic properties of their L1. Further, the use of a conceptual number was not modulated by L2 proficiency. This suggests that no matter the L1, the production of subject-verb agreement in L2 English—at least as it relates to producing agreement features on the verb—relies upon the same underlying mechanisms (Foote, 2010; Nicol & Greth, 2003); group differences in previous work may be due to proficiency level or methodological differences. Despite nativelike performance among both L2 groups in the integration of conceptual and grammatical number, both groups produced more

TABLE 4. Summary of bootstrapped CIs for L2 speaker miscellaneous error data

	Proportion Noun-Number Marking Errors		Proportion Other Miscellaneous Errors	
	Hedge's *g*	CI (low, high)	Hedge's *g*	CI (low, high)
Sing. Ref.; Sg.-Sg.	0.00	-0.55, 0.58	0.39	-0.19, 0.96
Dist. Ref.; Sg.-Sg.	0.14	-0.44, 0.68	0.06	-0.48, 0.63
Sing. Ref.; Sg.-Pl.	0.20	-0.37, 0.79	0.09	-0.47, 0.65
Dist. Ref.; Sg.-Pl.	0.60	0.03, 1.20	-0.19	-0.76, 0.38

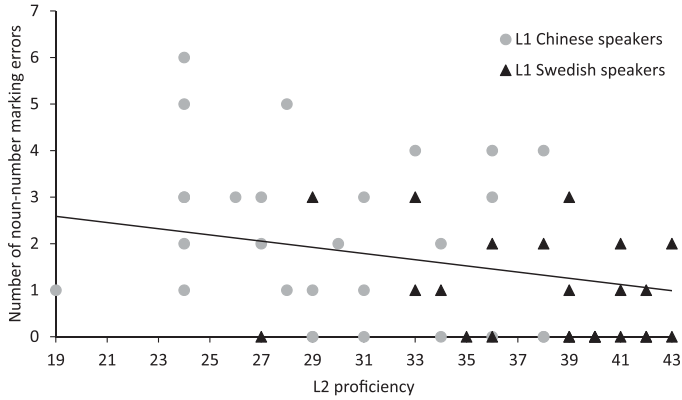

FIGURE 4. Correlation between L2 proficiency and noun-number marking errors on distributive referent items with a plural local noun.

agreement errors over the L1 baseline. This suggests a potential role of memory retrieval difficulties and cross-language interference at the point speakers must use the subject NP to compute number agreement on the verb and establish a controller-target dependency. This is consistent with recent proposals regarding the role of working memory and cross-language interference in L2 processing more generally (e.g., Cunnings, 2017).

In contrast to similar performance of the two L2 groups in subject-verb agreement, group differences emerged in noun-number marking errors during preamble repetition with conceptually plural items. This pattern was additionally modulated by L2 proficiency. The fact that L1 Swedish and L1 Chinese speakers with similar L2 proficiency exhibited similar error patterns suggests that while the absence of robust L1 NP number morphology may contribute to the difficulty of encoding the subject NP during L2 production, it cannot be the only factor. We propose that to acquire English plural morphology, L2 speakers must adopt thinking-for-speaking patterns that privilege inflectional morphology over other means of expressing number information, and that this process is difficult for all L2 speakers, not just for those with no equivalent L1 processing routines that require speakers to distinguish singular from plural meaning through inflectional morphology (Slobin, 1996; but see Jiang, Novokshanova, Masuda, & Wang, 2011). The present data suggest that this strategy shift poses lingering difficulty for L2 speakers, such that precisely those items that are conceptually but not grammatically plural elicit more noun-number marking errors. The influence of conceptual number on these errors suggests that less-proficient L2 English speakers, regardless of L1, have trouble encoding and retrieving grammatical information about the subject NP, making them susceptible to the presence of multiple possible syntactic plans and multiple conflicting informational cues.

CONCLUSION

The present study shows that like L1 English speakers, Swedish and Chinese L2 English speakers can integrate conceptual and grammatical number when producing agreement

in L2 English, despite subject-verb agreement not being a morphosyntactic feature of either L1. However, group- and L2 proficiency–related differences arose in number encoding, as reflected in number marking errors on subject NPs, highlighting how lower L2 proficiency—perhaps exacerbated by the absence of L1 noun number morphology—can still lead to increased difficulties associated with memory retrieval and cross-language interference when producing number marking in L2 English. By including a novel L2 population, Swedish L2 English speakers, and leveraging the distinction between grammatical and conceptual number and their diverse impacts on sentence production, the present study demonstrates how individual differences between speakers can asymmetrically impact subject-verb agreement and noun-number marking, while also underscoring that both processes are similarly influenced by grammatical and conceptual number information during L2 production. As such, this study provides a more nuanced understanding of how L2 proficiency, possibly in combination with cross-linguistic differences, can influence L2 agreement production.

NOTES

[1]Except for Foote (2010), few published L1 and L2 subject-verb agreement production studies report effect sizes. A power analysis was conducted using GPower 3.1 (Faul, Erdfelder, Lang, & Buchner, 2007) based on the effect size for the main effect of distributivity among late L2 learners from Foote (Cohen's $f = .65$), a significance level of $\alpha = .05$, and a desired power level of .9. This revealed that nine participants per group were necessary to detect a significant effect of distributivity in the present study. As the methodology used in Foote (2010) likely increased the magnitude of distributivity effects, we established a criterion of 24 participants per experimental group, comparable to other L2 studies (range: 15–38 participants; Hoshino et al., 2010, 2012; Nicol & Greth, 2003; Wei et al., 2015). This sample size would allow for detecting effects as small as $f = 0.35$.

[2]Three Swedish and four Chinese participants reported not learning English until school aged but provided no precise age at which they began English instruction. Therefore, the age of first exposure values in Table 1 are based on 21 Swedish and 24 Chinese participants. Four Chinese participants did not report how long they had lived in the United States. This information in Table 1 is based on data from 24 participants.

[3]Ideally the L1 Swedish and L1 Chinese speakers would be matched in L2 English proficiency. However, it would be difficult to achieve nonsignificant differences in proficiency between these two L2 speaker groups without extending recruitment to include populations beyond university-level student populations, which then risks introducing additional demographic differences between the two groups that could impact subject-verb agreement production. This would also lead to groups that are not representative of university-level student populations from these two L2 groups more generally.

[4]In instances of unequal variance, t-tests were run assuming unequal variance between populations. Corrected degrees of freedom and *p*-values are reported.

[5]Preliminary analyses revealed no effect of task order on participants' agreement production and the pattern of other significant effects remained identical to those reported here.

[6]A second analysis including miscellaneous responses that were still codable for verb-number marking revealed the same pattern of results.

[7]There were inadvertently 59 prescriptively singular items and only 41 prescriptively plural items across the experiment, including both experimental and filler items. To investigate whether there was a cumulative effect of hearing more prescriptively singular items as the experiment progressed, we ran an analysis including the number of prescriptively singular items previously heard as an additional fixed effect. The inclusion of the two-way interaction between this factor and group significantly improved the model fit ($\chi 2(2) = 6.62$, $p = .037$), because the L1 English speakers were less likely to produce plural agreement as the experiment progressed, but the proportion of plural responses did not vary as the experiment progressed for either L2 speaker group. Other significant effects remained identical to those reported in Table 3. As the primary goal of the present study is to investigate L2 agreement production and L2 production was unaffected by the number of singular items previously heard, this result will not be discussed further.

[8]A mixed-effect logistic regression model including L2 proficiency as an additional fixed-effect similarly revealed no significant effects or interactions with L2 proficiency on verb production.

REFERENCES

Antón-Méndez, I., & Hartsuiker, R. (2010). Morphophonological and conceptual effects on Dutch subject-verb agreement. *Language and Cognitive Processes*, *25*, 728–748.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*, 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Blattner, G. (2007). *Processing verbal arguments in a first and second language: The role of immersion experience* (Unpublished doctoral dissertation). Pennsylvania State University, University Park, PA.

Bock, K., & Cutting, J. C. (1992). Regulating mental energy: Performance units in language production. *Journal of Memory and Language*, *31*, 99–127.

Bock, K., & Miller, C. A. (1991). Broken agreement. *Cognitive Psychology*, *23*, 45–93.

Brehm, L., & Goldrick, M. (2016). *Competing plans explain agreement errors*. Poster presentation at the international workshop on language production. San Diego, CA.

Chen, L., Shu, H., Liu, Y., Zhao, J., & Li, P. (2007). ERP signatures of subject-verb agreement in L2 learning. *Bilingualism: Language and Cognition*, *10*, 161–174.

Christianson, K. (2016). When language comprehension goes wrong for the right reasons: Good-enough, underspecified or shallow language processing. *Quarterly Journal of Experimental Psychology*, *69*, 817–828.

Clahsen, H., & Felser, C. (2006). Grammatical processing in language learners. *Applied Psycholinguistics*, *27*, 3–42.

Cumming, G. (2014). The new statistics: Why and how. *Psychological Science*, *25*, 7–29.

Cunnings, I. (2017). Interference in native and non-native sentence processing. *Bilingualism: Language and Cognition*, *20*, 712–721.

Eberhard, K. M. (1999). The effect of conceptual number on the production of subject-verb agreement in English. *Journal of Memory and Language*, *41*, 560–578.

Eberhard, K. M., Cutting, J. C., & Bock, K. (2005). Making syntax of sense: Number agreement in sentence production. *Psychological Review*, *112*, 531–559.

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3. *Behavior Research Methods*, *39*, 175–191.

Foote, R. (2010). Age of acquisition and proficiency as factors in language production: Agreement in bilinguals. *Bilingualism: Language and Cognition*, *13*, 99–118.

Gibson, E., Bergen, L., & Piantadosi, S. T. (2013). Rational integration of noisy evidence and prior semantic expectations in sentence interpretation. *Proceedings of the National Academy of Sciences*, *110*, 8051–8056.

Gillespie, M., & Pearlmutter, N. J. (2011). Hierarchy and scope of planning in subject-verb agreement production. *Cognition*, *118*, 377–397.

Hartsuiker, R. J., Kolk, H. H. J., & Huinck, W. J. (1999). Agrammatic production of subject-verb agreement: The effect of conceptual number. *Brain and Language*, *69*, 119–160.

Haskell, T. R., Thornton, R., & MacDonald, M. C. (2010). Experience and grammatical agreement: Statistical learning shapes number agreement production. *Cognition*, *114*, 151–164.

Hatzidaki, A., Branigan, H. P., & Pickering, M. J. (2011). Co-activation of syntax in bilingual language production. *Cognitive Psychology*, *62*, 123–150.

Hoshino, N., Dussias, P. E., & Kroll, J. F. (2010). Processing subject-verb agreement in a second language depends on proficiency. *Bilingualism: Language and Cognition*, *13*, 87–98.

Hoshino, N., Kroll, J. F. K., & Dussias, P. E. (2012). Psycholinguistic perspectives on L2 speech production. In M. Sanz & I. M. Igoa (Eds.), *Applying language science to language pedagogy: Contributions of linguistics and psycholinguistics to second language teaching* (pp. 107–129). Newcastle, UK: Cambridge Scholars Publishing.

Humphreys, K. R., & Bock, J. K. (2005). Notional number agreement in English. *Psychonomic Bulletin and Review*, *12*, 689–695.

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

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological congruency and the acquisition of L2 morphemes. *Language Learning*, *61*, 940–967.

Kirby, K. N., & Gerlanc, D. (2013). BootES: An R package for bootstrap confidence intervals on effect sizes. *Behavior Research Methods*, *45*, 905–927.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Lorimor, H., Jackson, C. N., & Foote, R. (2015). How gender affects number: Cue-based retrieval in agreement production. *Language, Cognition and Neuroscience*, *30*, 947–954.

Lorimor, H., Bock, K., Zalkind, E., Sheyman, A., & Beard, R. (2008). Agreement and attraction in Russian. *Language and Cognitive Processes*, *23*, 769–799.

MacWhinney, B. (2012). The logic of the unified model. In S. Gass & A. Mackey (Eds.), *The Routledge handbook of second language acquisition* (pp. 211–227). Hoboken, NJ: Taylor and Francis.

Nicol, J., & Greth, D. (2003). Production of subject-verb agreement in Spanish as a second language. *Experimental Psychology*, *50*, 196–203.

Patson, N. D., & Husband, E. M. (2016). Misinterpretations in agreement and agreement attraction. *The Quarterly Journal of Experimental Psychology*, *69*, 950–971.

R Development Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from http://www.R-project.org.

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools.

Slevc, L. R., & Martin, R. C. (2015). Syntactic agreement attraction reflects working memory processes. *Journal of Cognitive Psychology*, *28*, 773–790.

Slobin, D. I. (1996). From "thought and language" to "thinking for speaking." In J. J. Gumperz & S. C. Levinson (Eds.), *Rethinking linguistic relativity* (pp. 70–96). Cambridge, UK: Cambridge University Press.

Thornton, R., & MacDonald, M. E. (2003). Plausibility and grammatical agreement. *Journal of Memory and Language*, *48*, 740–759.

Vigliocco, G., & Hartsuiker, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, *128*, 442–472.

Wei, X., Chen, B., Liang, L., & Dunlap, S. (2015). Native language influence on the distributive effect in producing second language subject-verb agreement. *Quarterly Journal of Experimental Psychology*, *68*, 2370–2383.