Thai Learners of English are Sensitive to Number-Agreement Violations

Teeranoot Siriwittayakorn
English Department, Faculty of Humanities
Chiang Mai University, 239 Huay Kaew Rd.,
Muang, Chiang Mai 50200 Thailand
teeranoot.s@cmu.ac.th

Edson T. Miyamoto
Center for Meta-Learning
Future University Hakodate, 116-2 Kameda
Nakano-cho, Hakodate, Japan
miyamoto@alum.mit.edu

Abstract

We report a reading-time experiment investigating how native Thai speakers process sentences with subject-verb number agreement in English as a second language. Participants were slower to read sentences containing agreement violations, in a manner similar to what has been reported for native English speakers. The results add to a growing literature according to which learners can acquire knowledge of number agreement even if their native language lacks it. This suggests that learners are not constrained by the features available in their native languages, and are able to acquire new features and put this knowledge to use when reading sentences for content.

1 Introduction

Following work on programming languages (Aho et al., 2007; and references therein), human sentence comprehension, or *parsing* broadly speaking, is often assumed to involve two components: a knowledge base (the grammar) and an algorithm that uses the knowledge base to process sentences (the parser). Moreover, it is commonly assumed that there is a single parser for all human languages; therefore, a child only needs to learn the grammar to be able to process sentences in a language (see Fodor, 1998, for detailed discussion, and on the impossibility for a child to learn the grammar and a language-specific parser at the same time).

A natural extension is that adults learning a second language (L2) only need to learn the grammar of the new language. The algorithm to use that knowledge is assumed to be the same as the parser for the learners' native language (L1). Therefore, behavioral differences between native speakers and L2 learners when processing sentences should be the result of differences in the knowledge base.

Within this framework, it is usually not enough to show that learners know some features of the L2 grammar. For example, in a traditional L2 task, we could ask learners of English to judge whether a sentence (e.g., *The keys is near the pencil*) is grammatical to determine whether they know that the subject and the verb have to agree in number in English. In this paper, we assume that most college students who have studied English know the basic rules of its number agreement system.

The more crucial question in this framework is to determine whether learners have acquired that knowledge and incorporated it to their L2 grammar, so that they can rapidly access it to process sentences in a manner that approaches L1 speakers' behavior. For this type of question, we can collect reading times to determine if learners slow down in situations in which native English speakers are known to be slow. For example, native English speakers are slow to read number-agreement violations as in (1) (Wagers et al., 2009; and references therein).

(1) The key to the cabinet are on the table.

33rd Pacific Asia Conference on Language, Information and Computation (PACLIC 33), pages 237-243, Hakodate, Japan, September 13-15, 2019

We report data on native Thai speakers to investigate how they process number agreement when reading L2 English. Thai does not have number markers or subject-verb agreement in general (see Iwasaki and Ingkaphirom, 2009, on Thai grammar). According to some early studies, learners are unable to keep track of subject-verb agreement in L2 when reading for comprehension if their L1 does not have that kind of relation (Hawkins and Chan, 1997; Jian, 2004; *inter alia*). Contrary to those claims, we suggest that speakers whose L1 does not have number agreement can display sensitivity to number-agreement violations in L2, extending previous results (Wen et al., 2010; Wilson and Miyamoto, 2015; *inter alia*).

2 Previous studies

According to previous literature, learners cannot acquire and rapidly manipulate features of L2 that are not available in their L1 (Hawkins and Chan, 1997; *inter alia*). For example, native speakers of languages that have number agreement (e.g., Russian) have been shown to be slow to read agreement violations in L2 English, similar to what has been reported for native English speakers; whereas native speakers of languages that do not have number agreement (e.g., Chinese, Japanese) do not show sensitivity to such violations (Jiang, 2004; Jiang et al., 2011).

However, there is an alternative way of interpreting these results. If we assume that L2 learning involves modifying L1 knowledge to approximate L2 (Schwartz and Sprouse, 1996), it may take longer for learners to acquire number agreement in L2 when their L1 lacks such feature. Moreover, the effects of individual variation (e.g., learners' proficiency) on language acquisition may be greater when learners acquire a feature from scratch, adding noise to experimental results. Therefore, it is conceivable that some previous studies (Jiang, 2004; Jiang et al., 2011; *inter alia*) although carefully conducted, failed to detect sensitivity to agreement violations because they did not take proficiency into consideration during the analyses.

In studies in which proficiency was included as a factor in the analyses, native speakers of Chinese and Japanese were shown to be sensitive to agreement violations while reading L2 English. One study used simple constructions involving agreement inside noun phrases (Wen et al., 2010; also, Yamada and Hirose, 2012, for data on more complex constructions). Another study used constructions as in (1) in which a prepositional phrase (*to the cabinet*) intervenes between the head noun and verb (Wilson and Miyamoto, 2015).

Given those previous results indicating the influence of proficiency, we included English proficiency scores when analyzing the Thai speakers' reading time data.

3 Experiment

3.1 Participants

Thirty-three native Thai speakers, undergraduate students at Chiang Mai University, volunteered to participate in the experiment. One participant's data were excluded because the participant did not follow the instructions. Results for the remaining 32 participants are reported.

All participants started learning English at the age of six or later, had never lived abroad for six months or longer, and were all majoring in English. Previous studies that did not detect sensitivity to agreement violations (Jiang, 2004; Jiang et al., 2011; *inter alia*) recruited L2 learners living in the United States. It is unlikely that English majors living in Thailand had more exposure to English than learners living in the United States, but this possibility is being addressed in on-going work.

3.2 Method

Stimuli: There were 16 pairs of test items, in which grammaticality was manipulated by modifying the number of the head noun in subject position, so that the head noun was plural in the grammatical condition and singular in the ungrammatical condition. The verb was always *were* (see (2) for an example pair; all stimuli were from Wilson and Miyamoto, 2015, with mistakes such as spelling corrected).

(2)

(a) Grammatical condition
The chickens in the oven were completely burned.

(b) Ungrammatical condition
The chicken in the oven were completely burned.

If learners are sensitive to number-agreement violations, they should be slow to read the verb *were* (or the word immediately thereafter) in the ungrammatical condition compared to the same word in the grammatical condition, as has been reported for native English speakers (Wagers et al., 2009, and references therein).

There are reasons to predict that no such a difference would be observed. First, because Thai does not have number agreement, native Thai speakers may not be able to acquire number agreement in L2 English (Hawkins and Chan, 1997) as has been reported for Chinese and Japanese speakers reading L2 English (Jiang, 2004; Jiang et al., 2011). Second, the intervening prepositional phrase (PP; e.g., *in the oven*) may make it too difficult for learners to keep track of the agreement relation between head noun and verb. In particular, learners may be unable to build the hierarchical structure in which the PP modifies the head noun, and instead build a shallower structure in which *oven* is the sister of *chicken(s)* (such a simplified structure would be compatible with Clahsen and Felser, 2006). In this case the verb may be associated with *oven*, instead of *chicken(s)*, therefore making both conditions equally acceptable.

However, in the same way as native Japanese speakers reading L2 English in a more recent study (Wilson and Miyamoto, 2015), it is conceivable that native Thai speakers are sensitive to number-agreement violations as well. In which case, the ungrammatical condition in (1b) should be read more slowly at the verb or later. This would suggest that Thai speakers acquire knowledge of number morphology and are able to use it in a manner that resembles native English speakers.

There were 48 filler sentences and 32 sentences from another experiment whose structure was similar to the test items to distract participants' attention away from the point of the experiment. All of these sentences were grammatical.

Procedure: Doug Rohde's Linger program was used to present sentences in a word-by-word non-cumulative self-paced reading procedure. The critical region (the verb *were*) was always region 6. Each participant saw eight grammatical sentences and eight ungrammatical sentences, and only one version of each pair of items. The test items were interspersed with 48 fillers and 32 items from another experiment in pseudo-random order so that two test items did not follow in succession. Each sentence was followed by a yes/no comprehension question. Feedback was provided when participants' answer was incorrect.

After the reading-time experiment participants answered a c-test questionnaire, in which they had to complete the second half of every other word in five texts (from Babaii and Shahri, 2010). Such questionnaires have been used in the past as an effective measure of proficiency in the analyses of reading times (Wen et al., 2010; Wilson and Miyamoto, 2015), and the scores have been reported to correlate well with more traditional measures such as the TOEFL-ITP (Wilson and Miyamoto, 2015).

Analysis: Analyses were performed on R version 3.5.0 (R Core Team, 2018). Only reading times from trials for which the comprehension question was answered correctly were included in the analyses. Initial trimming eliminated reading times below 100 ms and those above 5000 ms as they were unlikely to reflect reading-related latencies (Baayen, 2008, pp. 243-244, for discussion).

First, we report results from analysis of variance (ANOVA) using untransformed reading times to provide a comparison with previous studies (e.g., Jiang, 2004). Moreover, like in these earlier studies we did not include learners' proficiency as a factor in this initial analysis. Similar trends were observed when log-transformed reading times were used.

Second, we report results from mixed-effects models using log-transformed reading times and including proficiency (i.e., c-test scores) as a factor. Log-transformed reading times are usually used to decrease the influence of extreme values, and are appropriate for learners in this experiment as there may be some extremely long reading times (e.g., for unknown words). Similar trends were observed with untransformed reading times. After the initial 100 - 5000 ms trimming step, model-based trimming was conducted to eliminate data points beyond three standard-deviations, and the model was refit with the remaining data (Baayen, 2008; pp. 243-244). For each region, the trimming procedure eliminated no more than 3% of the data.

For all models, by-participants and by-items random intercepts were included. Whether a term was included as by-participants or by-items random slope was determined through backward selection (Bates et al., 2015). Numerical factors were centered to facilitate interpretation and improve convergence of the models.

Results not reported were not reliable (*p*s > .1).

3.3 Results

Proficiency (c-test scores): The average for the c-test scores was 77.63% (range 29 to 93, SD 13.96).

Question-Response Accuracy: For the test items and fillers, participants' comprehension performance was 77.23% or higher (mean 90%). For the test items, participants scored 81.25% or higher (mean 95.31%). There was no difference between the grammatical (94.53%) and the ungrammatical conditions (96.09%; $p = .385$).

Reading Times: We report results from ANOVA and mixed-effects models separately.

ANOVA: In region 2 (the head noun *chicken*), there was a trend for a grammaticality effect as the ungrammatical condition was faster than the grammatical condition, marginally in the by-subjects analysis and reliably in the by-items analysis (F_1 (1, 31) = 3.56, $p = .069$; F_2 (1, 15) = 5.99, $p = .027$). This replicates previous studies (for native speakers, see Lee and Cochran, 2000; Wager et al., 2009; and for learners, see Jiang, 2004; Wilson and Miyamoto, 2015). One possible reason for this difference is that the head noun in the grammatical condition was plural (*chickens*), therefore it was one character longer than the head noun in the ungrammatical condition.

In the critical region (region 6, the verb *were*), where the difference between the two conditions was predicted, there was no effect of grammaticality (F_1 (1, 31) = 2.35, $p = .135$; F_2 (1, 15) = 0.72, $p = .409$).

Previous studies often reported a reliable difference in the next region (for native English speakers: Pearlmutter et al., 1999; Jiang, 2004; Wagers et al., 2009). In our experiment, there was a trend for the ungrammatical condition to be slower than the grammatical condition in the by-subjects analysis but not in the by-items analysis (F_1 (1, 31) = 3.36, $p = .076$; F_2 (1, 15) = 1.41, $p = .254$).

Mixed-effects Models 1: Two types of analyses were conducted with mixed effects models. In the first type of analysis, log-transformed reading times to each region were analyzed as a function of *grammaticality* so as the results can be compared to those from the ANOVAs.

In region 2, there was an effect of grammaticality such that the ungrammatical condition was read faster than the grammatical condition ($\beta = -0.15$, $p < .001$).

At the critical region, there was no effect of grammaticality, but the numerical trend was for the ungrammatical condition to be slower than the grammatical condition ($\beta = 0.026$, $p = .412$).

In region 7, the ungrammatical condition was read significantly more slowly than the grammatical condition ($\beta = 0.06$, $p = .049$).

Mixed-effects Models 2: Proficiency is likely to be an important factor when analyzing reading times (Wen et al., 2010; Yamada and Hirose, 2012; Wilson and Miyamoto, 2015); therefore, the second type of mixed-effects models included grammaticality, c-test score, and their interaction as factors.

In region 1, there was an effect of c-test score ($\beta = -0.01$, $p < .001$) such that the higher their score was, the faster participants read. There was also an interaction between c-test and grammaticality ($\beta = 0.005$, $p = .007$). This interaction was unexpected because the word in this region was always the same (the article *the*). Participants may sometimes pause at random at the beginning of a sentence, or may be affected by the previous trial (e.g., they tend to slow down when they make a mistake answering the question in the previous trial).

Because of this spurious effect, for the remaining regions log-transformed reading times to region 1 were added as a covariate (analyses without the covariate revealed similar trends).

In region 2, the ungrammatical condition was faster than the grammatical condition ($\beta = -1.25$, $p < .001$). There was also an effect of c-test score as reading times were faster as the c-test score increased ($\beta = -7.56$, $p = .009$). Moreover, the covariate was reliable as reading times to region 1 were associated with slow reading times to region 2 (as indicated by the positive estimate, $\beta = 4.89$, $p < .001$).

In regions 3 and 4, there was a main effect of covariate (region 3, $\beta = 0.33$, $p < .001$; region 4, $\beta = 0.29$, $p < .001$).

In region 5, there was a main effect of c-test score such that the reading times were faster as the c-test score got higher (β = -1.13, p = .012). The effect of covariate was also reliable (β = 2.32, p < .001).

In region 6 (the critical region), there was an effect of covariate (β = 0.21, p < .001).

In region 7, the ungrammatical condition was reliably slower than the grammatical condition (β = 0.06, p = .043). There was a marginal effect of c-test score (β = -0.006, p = .062) suggesting that reading times got faster as the c-test score increased. Moreover, the covariate was reliable (β = 0.17, p < .001).

For all the later regions, there was a marginal effect of c-test score indicating faster reading times as the c-test score increased (β = -5.19, p = .053). The covariate was reliable (β = 1.54, p < .001).

4 Discussion

The results from mixed-effects models indicate that native Thai speakers are sensitive to agreement violations when reading L2 English, in line with previous results reported for native Japanese speakers (Wilson and Miyamoto, 2015). Results from ANOVAs were less clear cut as the effect of grammaticality was marginal in by-subjects and not reliable in by-items analyses. This is similar to a previous study with Chinese learners of English (Jiang, 2004), which detected no effect of grammaticality in by-subjects or in by-items ANOVAs using sentences comparable to those used in our experiment.

Mixed-effects models are increasingly common in the analyses of behavioral data as they have various advantages over traditional analyses such as ANOVAs (Baayen et al., 2008; Jaeger, 2008; and references therein). Our results (see the section titled Mixed-Effects Models 1) suggest that using mixed-effects models allow us to detect differences that were missed in previous studies, leading to rather different conclusions with respect to L2 acquisition and parsing.

More detailed mixed-effects analysis (see the section titled Mixed-Effects Models 2) indicates that proficiency (i.e., c-test score) contributes to explaining Thai learners' reading times as higher scores were associated with faster reading times. However, differences in proficiency did not affect the sensitivity to agreement violations as there was no interaction between grammaticality and proficiency. A previous study reported such an interaction suggesting that sensitivity to agreement violations were only found in native Japanese speakers with high proficiency in L2 English (Wilson and Miyamoto, 2015). However, preliminary analyses with mixed-effect models including the data for both Japanese and Thai participants revealed no 2-way interaction between L1 and grammaticality, or 3-way interaction between L1, grammaticality and proficiency, thus suggesting that the two groups' reading times are similar with respect to grammaticality.

The Thai participants' proficiency (mean c-test score: 71.16) was higher than the Japanese participants' (59.71). One possible reason for this difference is that the Thai participants were English majors. For a better comparison with the Japanese data, a new version of this study with Thai speakers not majoring in English is under way.

5 General Discussion

We reported the results of a reading-time experiment indicating that native Thai speakers are sensitive to number-agreement violations in L2 English. This suggests that not only can Thai speakers acquire an L2 feature absent in their L1, but they can also use this knowledge in a manner similar to that of native English speakers.

There is no overt morphological number marking in Thai; therefore, according to some past proposals (Hawkins and Chan, 1997), Thai learners of English should not be able to acquire agreement knowledge, contrary to our results.

According to other proposals (Clahsen and Felser, 2006), the PP intervening between the subject head noun and the verb in sentences as those in (2) may be too complex for learners to keep track of the agreement relation across it. Parsing such a complex structure may impose demands beyond learners' cognitive resources, forcing them to rely on simplified syntactic structures (as well as lexical information and world knowledge) to accomplish the task at hand. However, such a view is not easily reconciled with the agreement violation sensitivity detected in our experiment. Keeping track of agreement relations was unnecessary to interpret the sentences in our experiment. Nevertheless, that is what participants

seemed to have done without being aware of it (when asked after the experiment, participants never mentioned anything unusual or that some sentences were ungrammatical).

Our results are compatible with the view that learners start with the knowledge of their L1 and modify this knowledge to learn L2 (Schwartz and Sprouse, 1996). When the L1 does not contain agreement features, participants may take longer or may not be as consistent in acquiring the agreement system. This is not incompatible with past results in which learners displayed agreement sensitivity only if their L1 had agreement relations (Jiang et al., 2011).

Logically speaking, acquiring the knowledge is a necessary but not sufficient condition to guarantee that learners behave in a way similar to native English speakers. However, if we assume that the parser is universal to all human languages (Fodor, 1998), then acquiring the knowledge is indeed enough for us to expect L2 learners to approach native readers' behavior as was the case in our experiment.

The Thai speakers' data together with previously reported Japanese speakers' data (Wilson and Miyamoto, 2015) indicate that L2 learners can approach native speakers' performance even if their starting point (their L1) differs in crucial ways from the target L2.

This study is part of an on-going project to investigate L2 English parsing by native speakers of languages that for the most part lack number morphology, namely, Chinese, Japanese and Thai. Despite previous proposals and results that claimed otherwise (Hawkins and Chan, 1997; Jiang, 2004; Jiang et al., 2011), preliminary results indicate that speakers of these languages can display sensitivity to agreement violations. One first goal is to investigate in detail to what extent learners approach native speakers' way of processing agreement. There are detailed results on how native English speakers process number morphemes and the situations in which *intervention effects* can occur (Wagers et al., 2009). Our prediction is that similar trends would be observed with speakers of Chinese, Japanese and Thai reading L2 English.

Another goal is to investigate how L1 affects L2 learning. According to Corder (1981), similarities between L1 and L2 can facilitate acquisition by decreasing the steps in the learning process. Although all three languages lack number agreement, they differ in how similar they are to English in other respects such as word order. For example, (a) Thai and Chinese have Subject-Verb-Object (SVO) word order like English, whereas Japanese is SOV; (b) relative clauses in Thai, like in English, are postnominal (they follow the modified noun), whereas relative clauses are prenominal in Chinese and Japanese; (c) as in English, adjectives precede the head noun in Chinese and Japanese, but in Thai adjectives follow the head noun. From (a) – (c), Thai and Chinese are more similar to English than Japanese is. The question then is whether this type of similarity metric would have an impact on how learners can acquire number agreement in English. Factors such as motivation and attitude towards English are also been measured to eliminate some basic confounds.

6 Conclusion

The present study investigated the processing of English number agreement by Thai learners. The results show that similar to native English speakers, Thai learners slowed down when encountering an agreement violation. This result together with other recent studies (Wen et al., 2010; Yamada and Hirose, 2012; Wilson and Miyamoto, 2015) indicate that the ability for learners to acquire L2 knowledge is not restricted by the knowledge of their L1. Learners whose L1 lacks some crucial aspects of L2 are still capable of acquiring such missing knowledge.

Acknowledgment

This research was partly supported by MEXT/JSPS KAKENHI Grant Number 19K00609 to the second author.

References

Aho, A. V., Lam, M. S., Sethi, R., & Ullman, J. D. (2007). *Compilers: principles, techniques, and tools* (2 ed.). USA: Pearson Education.

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

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of memory and language, 59*(4), 390-412.

Babaii, E., & Shahri, S. (2010). Psychometric rivalry: The C-test and the close test interacting with test takers' characteristics. In R. Grotjahn (Ed.), *The C-Test: Contributions from Currect Research* (pp. 41-56). Frankfurt:Lang.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *arXiv preprint arXiv:1506.04967*.

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

Corder, S.P. (1981). *Error analysis and interlanguage*. New York: Oxford University Press.

Fodor, J. D. (1998). Learning to parse? *Journal of Psycholinguistic Research, 27*(2), 285-319.

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.

Iwasaki, S., & Ingkaphirom, P. (2009). *A reference grammar of Thai*: Cambridge University Press.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of memory and language, 59*(4), 434-446.

Jiang, N. (2004). Morphological insensitivity in second language processing. *Applied Psycholinguistics, 25*, 605-634.

Jiang, N., Novokshanova, E., Masuda, K., & Wang, X. (2011). Morphological Congruency and the acquisition of L2 morphemes. *Language Learning, 61*(3), 940-967.

Lee, C. H., & Cochran, M. F. (2000). Controlling two confounding variables in word length: "Vanished word-length effect". *Reading Psycholgy, 21*, 57-66.

Pearlmutter, N. J., Garnsey, S. M., & Bock, K. (1999). Agreement processes in sentence comprehension. *Journal of memory and language, 41*, 427-456.

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Schwartz, B. D., & Sprouse, R. A. (1996). L2 cognitive states and the Full Transfer / Full Access model. *Second Language Research, 12*, 40-72.

Wagers, M. W., Lau, E. F., & Phillips, C. (2009). Agreement attraction in comprehension: Represetation and processes. *Journal of memory and language, 61*, 206-237.

Wen, Z., Miyao, M., Takeda, A., Chu, W., & Schwartz, B. D. (2010). *Proficienc effects and distance effects in nonnative processing of English number agreement.* Paper presented at the Boston University Conference on Language Development, Boston, USA.

Wilson, B. G., & Miyamoto, E. T. (2015). *Proficiency effects in L2 processing of English number agreement across structurally complex material.* Paper presented at the the 151st Meeting of the Linguistic Society of Japan, Nagoya University.

Yamada, T., & Hirose, Y. (2012). *Singular / plural asymmetry in Japanese EFL learners' sensitivity to English number dis/agreement.* Paper presented at the Japanese Cognitive Science Society, 29.