THE JOURNAL OF ASIA TEFL
Vol. 19, No. 3, Fall, 2022, 896-910
http://dx.doi.org/10.18823/asiatefl.2022.19.3.9.896

The Journal of Asia TEFL

http://journal.asiatefl.org/

e-ISSN 2466-1511

The Modulating Role of Mean Dependency Distance in the Backward Transfer of L3 on the Interpretation of L2 Reflexives

Hee-Don Ahn
Konkuk University, Korea

Yongjoon Cho
Konkuk University, Korea

Jong-Bai Hwang
Konkuk University, Korea

Jung Hyun Lim
Konkuk University, Korea

Cui Mao
Shandong University of Technology, China

Based on the supportive evidence for the existence of backward transfer of L3 on the interpretation of L2 reflexives detected by Ahn & Mao (2019), this paper explores potential factors that possibly affect the process of backward transfer of L3. Successful interpretation of reflexives--the searching for the antecedents to which the reflexives might be bound--is subject to the constraints of syntactic and discoursal factors, given that English is more syntactic-oriented, while Chinese and Korean are more pragmatic-oriented. The present study compares the accuracy rates in the Truth Value Judgment Task to examine the role of syntactic complexity in Chinese-English (CE), Korean-English (KE), and Chinese-English-Korean (CEK) language configurations. The results show that the Mean Dependency Distance (MDD) is a negative and significant predictor of the probability of accurate response in the TVJT. However, when the regression model is fitted by group, it is revealed that the regression coefficient of the MDD strikes a significant level with the CE and the CEK groups, but not with the KE group. The CEK group is more capable of rejecting the influence of syntactic complexity than the CE group, either because of their enhanced meta-linguistic knowledge or because of weakened reliance on syntactic constraints. The findings are in line with both the Cumulative Enhancement Model and the Foreign Language Effect Model.

Keywords: reflexive, L3 acquisition, backward transfer, Mean Dependency Distance

Introduction

Successful comprehension of reflexives involves the ability to link reflexives and their antecedents quickly and accurately under the constraints of syntactic and discoursal factors. A great number of studies

have focused on exploring how syntactic constraints influence antecedent retrieval during processing (Badecker & Straub, 2002; Clackson et al., 2011; Cunnings & Felser, 2013; Xiang et al., 2009). On the other hand, there are studies to show that both syntactic and pragmatic computations are involved in the interpretation of anaphors (Lee, 2008; Huang, 1991, 2000a, 2000b, 2004, 2013, 2016; Sperlich, 2016).

Huang (2000, 2013, 2016) explored intra-sentential anaphors and concluded that languages such as English, French, and German were syntactic-oriented, whereas languages like Chinese, Japanese, and Korean were more pragmatic-oriented. Based on the binary classification of world languages, we assume that language learners with their L1, L2 and/or L3 showing different orientations may be susceptible to the interplay of the syntactic and the pragmatic constraints in the acquisition of reflexive bindings. Therefore, the present study employed a story-based Truth Value Judgment Task (TVJT) with participants from the CE, the CEK, and the KE groups. Using the Mean Dependency Distance (MDD) of the stimuli as a matrix denoting syntactic complexity of the target sentences, the study probed into the potential role of the MDDs in the accuracy of reflexive resolutions in L2 English and explored whether the pattern of the backward transfer with the L3 group would be modulated by sentence complexity.

Literature Review

Mean Dependency Distance (MDD)

The notion of syntactic dependency is widely accepted in computational linguistics (Nivre 2006) and theoretical linguistics (Hudson, 2007). It suggests that the syntactic structure of a sentence consists of nothing but dependencies between individual words. The syntactic dependency relation is binary between two linguistic units, and asymmetrical with one unit being the governor and the other being the dependent (Husdson, 2007).

Dependency distance (Liu, 2008) is a dynamic measurement of the linear distance or linear positional difference intervening between two words that are syntactically related within a sentence (Hudson, 2010). It is generally taken as a matrix of syntactic complexity in linguistics and is widely used in cognitive science as an index of working memory load (Liu, 2008, 2010). Dependency distance is found to be positively related to syntactic complexity, and thus the increment in dependency distance poses a heavier burden on working memory.

Studies On MDD

Gibson (1998) proposed the Syntactic Prediction Locality Theory, according to which locality exerted substantial influence on memory and integration costs. An increase in the distance between the dependent word and its head would result in a greater integration cost.

Liu (2008) examined twenty corpora from different languages with syntactic annotation of dependency and found that the MDD of the twenty languages fell into the range of 1.798-3.662, which indicated that the human language parser preferred a minimized distribution of dependency distance. It was assumed that the increment of MDD was positively related to the increase in the number of words that had to be kept in memory in parsing the sentence, posing a greater burden on the working memory capacity.

Jiang and Liu (2015) made comparisons on the cross-linguistic variation in the co-relations between sentence length and dependency distance and concluded that the MDDs of Chinese sentences were always higher than those of English sentences, and the MDDs of sentences were positively related to the sentence length.

Jiang and Ouyang (2017) applied the perspective of MDD to the field of L2A and examined the syntactic development of Chinese-English learners from 8 grade. They calculated the MDDs of the learners' compositions and found that there was a significant increase in the MDDs from grade one in

junior high schools to grade two in universities, which indicated that the sentence complexity of the learners' compositions augmented with the increase of learners' grade, thus the increase in proficiency presumably.

Chow and Zhou (2018) conducted an eye-tracking experiment to investigate the relationship between the filler-gap dependency distance and active gap-filling in real-time sentence processing. The results showed supportive evidence for active gap-filling when the dependencies were grammatically licensed, but the effect tended to be negatively related to the increase in the dependency distance.

Jiang and Jiang (2019) explored the relationship between MDD and cognitive load, taking non-fluency in interpreting tasks as a behavioural measure. In their study, the effect of the maximum dependency distance of the original text on the translator's non-fluency, and the moderating effect of the syntactic structure were examined through an English-Chinese visual translation experiment. The results showed that the non-fluency of the sentences with longer maximum dependency distance was significantly higher than those with shorter maximum dependency distance.

Methodology and Procedures

The present study is a follow-up study of Ahn and Mao (2019), which examined the reverse transfer of L3 Korean on L2 English in the interpretation of reflexive binding with Chinese and Korean participants. In what follows, we outline the core experimental design in the paper which we adopted here as relevant background for the current experiments and statistical analyses.

Participants

Fifty-three native Chinese undergraduates (12 males, 41 females; average age = 19) were recruited and tested at a university in Shandong Province of China. All participants reported starting to learn English in primary school (Average Age of Acquisition = 8.69). None of the participants reported having any experience learning English outside the classroom setting in China. Thirty of them reported having no prior experience with the Korean language (CE), and twenty-three were in their first or second year of learning Korean (CEK, average Korean AOA = 18.08). Monetary compensation was paid for their participation upon completion. Twenty native Korean undergraduates were recruited and tested at a university in Seoul, Korea (3 males, 17 females; freshmen). The reported average AOA of English is 8.26. None of them reported having any experience learning English outside the classroom setting in Korea, nor any experience of learning a third language (KE).

Materials and Procedures

All three groups completed a Cambridge Quick Placement Test, a story-based Truth Value Judgment Task (TVJT), as well as a questionnaire on the linguistic background. All tests were off-line and were taken in classroom settings. Instructions for the three tests were provided in English. Participants were asked to complete the items as fast and accurately as possible.

The Cambridge Quick Placement Test

The Cambridge Quick Placement Test (Version 2) was adopted to assess the English proficiency of the participants. The test consisted of 60 multiple-choice questions targeted at reading and grammar abilities. The test paper was printed out and distributed to participants in a classroom setting. There was no time limit by which the participants had to complete the test, and the participants were instructed to answer all the questions as fast and accurately as possible. Most participants could complete the proficiency test in approximately 35 minutes.

Participants in the present study were placed on the B level (intermediate: 30-48) based on the scores in the placement (1-60).

The Truth Value Judgment Task

English anaphors allow only for short-distance binding, while Chinese and Korean anaphors can be both short- and long-distance bound.

(1) a. John$_i$ thought Mary$_j$ trusted himself*$_{i/j}$.
b. John$_i$ renwei Mary$_j$ xiangxin ziji$_{i/j}$.
John think Mary trust self
'John thought that Mary trusted self.'
c. John$_i$-i Mary$_j$-ka caki$_{i/j}$-lul sinloyhayssta-ko sayngkakhayssta.
John-Nom Mary-Nom self-Acc trusted-Comp thought
'John thought that Mary trusted self.'

In (1a), the English reflexive *himself* must take the embedded subject *Mary*, but not the matrix subject *John* as its antecedent. In (1b) and (1c), on the contrary, the Chinese *ziji* and the Korean *caki* can take the matrix subject as well as the embedded subject as their antecedent. Thus, anaphors in English must be locally bound, while Chinese and Korean anaphors can be long-distance bound.

However, Pollard and Sag (1992,1994) indicated that there are two types of anaphors in English: 'normal' or 'true' anaphors and 'exempt' anaphors. A normal anaphor is illustrated in (1a), which is subject to short-distance binding, whereas 'exempt' anaphors marginally permit long-distance bound, as shown in the following examples (here we demonstrate only a few) in (2):

(2) a. Bill said that the rain had damaged pictures of himself.
b. Max boasted that the queen invited Lucie and himself for a drink.
c. Joe worried that his girlfriend was pulling away from himself.
d. John decided that Mary's remarks had been intended for himself.
e. Tom thinks that Julie admires everyone but himself.

Notice that reflexives in these sentences may refer to the antecedents across the clause boundary. Accordingly, these 'exempt' reflexives in English sharply contrast with normal (true) reflexives in that they can be long-distance bound. We will not attempt to account for the nature of the two types of anaphors in English since it is not our direct concern and is beyond the scope of the present research. This distinction is employed for designing the experimental items for our methodological purposes only.

The story-based Truth Value Judgment Task was adapted from the one used in Yoshimura, Nakayama, Sawasaki, Fujimori and H. Shimizu. (2012). Participants read a short dialogue of 3 to 5 turns between two interlocutors first and were asked to judge if the corresponding sentence is true or false according to what they had just read. Both the narratives and the testing sentences were written in English.

A total of 38 different narratives with different test sentences were presented on the printed paper to participants. Before the test session, there were 2 practice items. The test session contained 12 fillers and 24 test items, with the testing items being balanced in the binding distance, finiteness, and truth value (See Table 1 for the balancing of test items).

TABLE 1
The Balancing of Test Items

Binding Type	Finiteness	Truth Value	Number
Normal Anaphor (Short Distance Binding)	Finite	True	3
		False	3
	Non-Finite	True	3
		False	3
Exempt Anaphor (Long Distance Binding)	Finite	True	3
		False	3
	Non-Finite	True	3
		False	3
Fillers		True	6
		False	6

The test sentences were designed in such a way that only one interpretation of binding was allowed, and the truth value of the sentence was determined by the narratives (see below for examples).

Example 1. Short-distance, Non-finite, False, Normal Anaphor
Dave: Could you do me a favor, Ralph?
Ralph: Sure.
Dave: A very important customer is coming tomorrow, so don't forget to contact me if you don't know where you should show him around. O.K.?
Ralph: O.K.

[Sentence] Dave advised Ralph to talk to himself. YES NO

Example 2. Long-distance, Finite, True, Exempt Anaphor
Mary: Hi, Sue. I read your boyfriend's poems. They are so beautiful.
Sue: Well, he wrote superb poems, but not a single one dedicated to me.
Mary: He will.

[Sentence] Mary believes that Sue's beloved boyfriend will write a poem dedicated to herself. YES NO

The sequential order of the 36 items in the test session was randomized into 10 different versions (Test Paper 1-10), and the participant got one of the test papers randomly. The participants accomplished the same two practice items before they continued to do the test session. There was no time limit by which the participants had to complete the test. The time the participants spent on this test ranged from 25 to 40 minutes.

Upon the accomplishment of the test session, participants were asked to complete a questionnaire on their linguistic background.

Questionnaire

The linguistic background questionnaire was written in English and contained 4 parts with 18 short answer and multiple-choice questions, as well as such personal information as name, class, and student ID number. The primary objective of the questionnaire was to obtain a deeper understanding of the participants' linguistic and educational backgrounds. Part 1 with its 4 short questions is on participants' L1 and the dominant language at home. Part 2 involves 2 short questions and 3 multiple-choice questions,

which are targeted at participants' L2 learning experience. Part 3, slightly different from part 2, was concentrated on the AOA and frequency of use of L3. If participants did not have the experience of learning an L3, they could skip to part 4, which asked about the experience of working, studying, or living abroad for more than 3 months. This part had 2 short questions and 2 multiple-choice questions. The questionnaire took about 5 minutes to be completed.

Procedures for Calculating MDD

According to Oya (2011), there are three steps in calculating the MDD of a sentence. Firstly, put the target sentence into the Stanford Parser, then the output will be presented in a list of triples. Each triple involves the syntactic relation between the governor and the dependent, followed by the governor along with its linear position in the target sentence, and the dependent with its linear position. The output of the parsing results is manifested in the following format as demonstrated in Figure 1:

Susan hopes Linda's advisor will write a letter of recommendation for herself.
nsubj(hopes-2, Susan-1)
root(ROOT-0, hopes-2)
nmod:poss(advisor-5, Linda-3)
case(Linda-3, 's-4)
nsubj(write-7, advisor-5)
aux(write-7, will-6)
ccomp(hopes-2, write-7)
det(letter-9, a-8)
dobj(write-7, letter-9)
case(recommendation-11, of-10)
nmod:of(letter-9, recommendation-11)
case(herself-13, for-12)
nmod:for(write-7, herself-13)
Tokens: 13

Figure 1. The results of the Stanford parser.

The first line in Figure 1 is the sentence that has been put into the Stanford Parser. The triple "*nsubj(hopes-2, Susan-1)*" indicates that the second-word *hopes* in the example sentence has a dependent word *Susan*, which is the first word in the sentence. The dependency type between these two words is *nsubj*, that is, nominal subject. The last line in Figure 1 tells that there is a total number of 13-word tokens in the example sentence. Table 2 summarizes the dependency relations between word pairs in the sample sentence.

TABLE 2
Dependency Relations of Example 1

Relation	Governor	Position of the governor	Dependent	Position of the dependent	Distance
root	ROOT	0	hopes	2	2
nsubj	hopes	2	Susan	1	1
nmod:poss	advisor	5	Linda	3	2
case	Linda	3	's	4	1
nsubj	write	7	advisor	5	2
aux	write	7	will	6	1
ccomp	hopes	2	write	7	5
det	letter	9	a	8	1
dobj	write	7	letter	9	2
case	recommendation	11	of	10	1

nmod:of	letter	9	recommendation	11	2
case	herself	13	for	12	1
nmod:for	write	7	herself	13	6

Secondly, calculate the MDD from the parser output. As is shown in Table 2, the format of the Stanford Parser output makes it easy to calculate the MDD. For example, in the relationship of "*nsubj(hopes-2, Susan-1)*", the distance of the dependency between the governor *hopes* and the dependent *Susan* is 2-1 = 1. As Oya (2011) illustrated, the MDD of a sentence is the sum of the distance of all the dependencies in the sentence divided by the number of the dependencies of the sentence. It should be noted that the dependency relations, such as *punct* and *root*, are excluded in the calculation (Liu, 2008), where *punct* denotes punctuations and *root* is the main verb (which serves no governor, and therefore its dependency distance is defined as zero) in the sentence. Therefore, the way of calculating the MDD of a sentence can be presented in the following formula ((Liu, 2008):

Formula 1

$$\mathrm{MDD} = \frac{\sum_{i=1}^{n} |DD_i|}{n}$$

Reading from Formula 1, *n* indicates the total number of dependency relationships in the target sentence, and $|DD_i|$ is the absolute value of the dependency distance of the *i*th pair of words in the dependency relationship. Calculating the MDD of Example 1 in Formula 1, we can get the results as follows:

$$\text{MDD (Example 1)} = \frac{|1+2+1+2+1+5+1+2+1+2+1+6|}{12} = 2.0833$$

Therefore, the MDD of Example 1 is 2.0833. The same procedure is exploited to get the MDD of all the stimuli in the SD and LD binding conditions. The results are demonstrated in the following Table 3.

TABLE 3
MDD of the Stimuli

Stimuli ID	Binding Condition	Sum of |DD|	Num. of Dependencies	MDD
1	LD	15	8	1.875
2	LD	23	13	1.769
3	LD	24	11	2.182
4	LD	24	12	2.000
5	LD	26	12	2.167
6	LD	21	11	1.909
7	LD	25	12	2.083
8	LD	16	10	1.600
9	LD	21	11	1.909
10	LD	20	10	2.000
11	LD	29	13	2.231
12	LD	27	13	2.077
13	SD	8	5	1.600
14	SD	8	5	1.600
15	SD	17	7	2.429
16	SD	11	7	1.571
17	SD	10	7	1.429
18	SD	15	8	1.875
19	SD	10	6	1.667
20	SD	8	5	1.600
21	SD	17	7	2.429
22	SD	11	7	1.571

23	SD	19	10	1.900
24	SD	9	6	1.500

The third step is to conduct statistical analysis with the MDD data as a matrix of sentence complexity. The following are the results of data analysis regarding MDD.

Results of MDD Analysis

MDD in SD and LD Binding Conditions

Total 73 participants made their judgment for the truth value on 24 items in the TVJT, which generated a total of 1752 cases. There were four missing values in the accuracy judgment, and thus 1748 cases were finally entered into the subsequent data analysis. Table 4 reported the descriptive data on the cases regarding the MDD, the binding condition, and the response accuracy [1].

TABLE 4
Case Processing Summary

Condition	Accuracy	Mean MDD	Std. Deviation	N
LD	Incorrect	1.9982	.15237	318
	Correct	1.9750	.18732	556
	Total	1.9835	.17568	874
SD	Incorrect	1.8526	.36810	288
	Correct	1.7194	.29017	586
	Total	1.7633	.32388	874
Total	Incorrect	1.9290	.28589	606
	Correct	1.8438	.27673	1142
	Total	1.8734	.28278	1748

An Independent-samples *t*-test with Binding Condition (SD vs. LD) as the independent variable and MDD as the dependent variable was conducted to explore if there was a statistically meaningful difference in the MDD between stimuli with normal anaphors and those with exempt anaphors (Figure 2).

[1] As an index of discrimination accuracy, D-prime scores (D′s) were calculated based on the proportion of "hits" obtained for each contrast and the proportion of "false alarms" (Macmillan and Creelman, 1991).

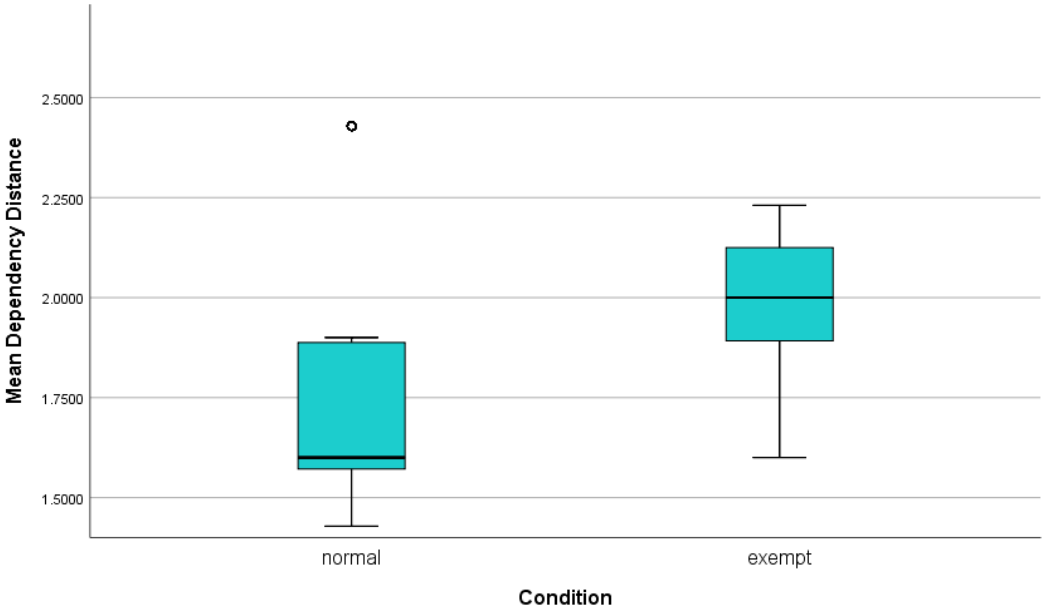

Figure 2. The Mean Dependency Distances in normal and exempt anaphors

As revealed in Figure 2, there was a significant difference in the means of the MDD between SD and LD binding conditions ($t(1750) = 17.603$, $p = .000$). The mean difference in the MDD (Mean Difference = -.2193) between the SD ($M = 1.7642$, $SD = 0.324$) and the LD ($M = 1.9835$, $SD = 0.175$) binding conditions were significant, with the former being substantially lower than the latter.

Variance in the Accuracy in TVJT

Given that the MDD was regarded as a matrix of syntactic complexity of sentences, and that the mean of MDD was found to be significantly lower in the SD than in the LD binding condition, it was plausible to hypothesize that the accuracy rate in the TVJT might be accounted for by the difference in the means of MDD between the two binding conditions. In the following part, a binary logistic regression model was fitted, with Accuracy (Correct vs Incorrect) as the dependent variable, and the MDD and the Binding Condition (Normal vs. Exempt) as two independent variables.

TABLE 5
Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	36.451	2	.000
	Block	36.451	2	.000
	Model	36.451	2	.000

As shown in Table 5, the regression model involving the MDD and the Binding Condition was statistically different from the null model ($\chi^2(2) = 36.451$, $p < .00$), which explained 2.8% (Nagelkerke R^2) of the variance in response accuracy in TVJT and correctly classified 66.1% of the cases. There was an improvement in the regression model compared to the 65.8% correct classification in the null model.

MDD was a negative and significant predictor of the probability of making correct judgment in the TVTJ (b = -1.118, s.e. = .192, Wald $\chi^2 = 34.043$, $p < .00$) (Table 6). Increasing MDD was associated with a reduction in the likelihood of making an accurate judgment in the TVJT, with the odds ratio (OR) (OR = 0.327, 95% CI OR: 0.225 - 0.476) indicating that for every one-unit increase on the MDD the odds of making an accurate judgment of reflexive binding changed by a factor of 0.327 (Table 6). Condition was

a non-significant predictor of making accurate judgment in the TVJT (b = .081, s.e. = .109, Wald χ^2 = .561, p = .454).

TABLE 6
Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1[a]	MDD	-1.118	.192	34.043	1	.000	.327	.225	.476
	Condition(1)	.081	.109	.561	1	.454	1.085	.877	1.342
	Constant	2.700	.351	59.120	1	.000	14.881		

a. Variable(s) entered on step 1: MDD, Condition.

Group Variance in the Accuracy in TVJT

Taken as a whole, the participants in the present experiment were shown to be susceptible to the variance of sentence complexity in their judgment in the TVJT. Since the CE, the CEK, and the KE groups manifested different patterns in their D-Prime scores in the TVJT in Ahn & Mao (2019), it was sensible to split the data by group in the present experiment and explore if there was group variance in the effect size of MDD as a predictor for response accuracy. To address this question, three binary logistic regression models were fitted with the three language groups respectively, using the MDD and the Binding Condition as predictors in the regression models. The results of the three groups were summarized in Table 7.

Table 7
Omnibus Tests of Model Coefficients with CE, CEK, and KE

		Chi-square	df	Sig.
CE	Step	28.615	2	.000
Step 1	Block	28.615	2	.000
	Model	28.615	2	.000
CEK	Step	9.506	2	.009
Step 1	Block	9.506	2	.009
	Model	9.506	2	.009
KE	Step	3.349	2	.187
Step 1	Block	3.349	2	.187
	Model	3.349	2	.187

As shown in Table 7, there existed group variance in the predictors that contributed to the regression model. The regression model involving MDD and Condition was significantly different from the null model with the CE group ($\chi^2(2) = 28.615$, p = .00), and the CEK group ($\chi^2(2) = 9.506$, p =.009), but not with the KE group ($\chi^2(2) = 3.349$, p = .187).

With the CE group, the regression model explained 5.3% (Nagelkerke R^2) of the variance in response accuracy in TVJT, and correctly classified 64.9% of the cases, compared to the 62.1% of correct classification in the null model. With the CEK group, the regression model accounted for 2.3% (Nagelkerke R^2) of the variance in response accuracy in TVJT. Nevertheless, the proportion of correct classification in the regression model remained the same at 64.7% as in the null model. With the KE group, the regression model with MDD and Condition only explained 1% of the variance in the accuracy of the TVJT. Like the CEK group, there was no salient improvement in the proportion of correct classification, remaining the same at 70.9% as the null model.

Looking at the two predictors entering the regression model (Table 8), MDD and Condition were detected to play different roles in the three groups. The increment in MDD was generally associated with a reduction in the likelihood of making accurate judgment in the TVJT with all the three groups, but it was revealed to be a significant and negative predictor in the accuracy of TVJT with the CE group (Wald χ^2 = 27.197, b = -1.576, s.e. = .302, p < .00) and the CEK group (Wald χ^2 = 7.242, b = -.913, s.e. = .339, p < .05), but not with the KE group (Wald χ^2 = 2.973, b = -.656, s.e. = .381, p > .05). The odds ratio

(OR) for MDD was 0.207 (95% CI OR: 0.114 - 0.374), 0.401 (95% CI OR: 0.206 - 0.780), and 0.519 (95% CI OR: 0.246 - 1.094) with the CE, the CEK and the KE group respectively. Condition with a *p-value* greater than 0.05 was not a significant predictor of the accuracy of TVJT with all three groups at the 5% level.

TABLE 8
Variables in the Equation with CE, CEK, and KE

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
CE	MDD	-1.576	.302	27.197	1	.000	.207	.114	.374
	Condition(1)	.253	.170	2.224	1	.136	1.288	.924	1.796
	Constant	3.336	.551	36.727	1	.000	28.113		
CEK	MDD	-.913	.339	7.242	1	.007	.401	.206	.780
	Condition(1)	-.080	.192	.175	1	.676	.923	.633	1.345
	Constant	2.367	.622	14.495	1	.000	10.670		
KE	MDD	-.656	.381	2.973	1	.085	.519	.246	1.094
	Condition(1)	.009	.216	.002	1	.967	1.009	.660	1.542
	Constant	2.123	.696	9.298	1	.002	8.354		

Discussion

The present experiment made a preliminary trial in searching for the potential influence of sentence complexity in the comprehension of English reflexives with L2 and L3 learners. Binary logistic regressions were administered by group, with the Mean Dependency Distance, a matrix denoting syntactic complexity, as a predictor for the accuracy in the TVJT. The analyses with the MDD yielded one major finding concerned with the predictive power of the MDD for response accuracy in the TVJT, and the group variance in the magnitude of the effect.

When analyzed as a whole, MDD was detected as a negative and significant predictor of the probability of accurate response in the TVJT. But when the regression model was fitted by group, it was revealed that the regression coefficient of the MDD struck a significant level with the CE and the CEK groups, but not with the KE group. With the CE and the CEK groups, the negative coefficient of the MDD could be taken as an indicator that the increase in MDD led to a reduction in the likelihood of making an accurate response in the TVJT.

This finding led to the assumption that the CE and the CEK groups were subject to the constraint of syntactic complexity in reflexive resolution, whereas there was no salient evidence showing the KE group was influenced by the MDD.

One possible explanation for this finding was that the L2 syntax of the CE and the CEK groups was less stable compared to that of the KE group due to their lower L2 proficiency. The augment in the MDD of the stimuli posed greater difficulties in processing the stimuli sentences for the CE and the CEK, resulting in a higher likelihood of failing to reject the syntactically wrong but contextually correct antecedents. By 'contextually correct', we meant to refer to the situation where the given context in the narration favored of one specific NP as the antecedent, whereas binding the reflexive to the favored NP in the stimulus sentence was syntactically incorrect. This explanation was in line with what Clahsen and Felser (2006) proposed as the Shallow Structure Hypothesis (SSH), which postulated that the syntactic representations computed by L2 learners during comprehension were shallower and less detailed than those of the native speakers. The SSH posits that L2 learners are less likely to make use of syntactic information in sentence processing; rather, they tend to rely more on lexical-semantic and pragmatic information, as well as world knowledge. We made an analogy that the L2 syntactic constructions of the CE and the CEK were weaker. With the increment in the syntactic complexity of the stimuli sentences, the participants were more succinct to the contextual influence of the narration, which lead to a syntactically incorrect antecedent for the reflexives in the stimuli sentences, resulting in a lower score in

the TVJT. In the case of the KEs who scored significantly higher in their L2 proficiency test, the L2 syntax might be stronger enough to guarantee a better job in overriding the contextual influence when the contextually favored antecedents were syntactically incorrect.

Looking at the negative regression coefficients and the odds ratio for MDD with the CE group (b = -1.576, OR = 0.207) and the CEK (b = -.913, OR = 0.401) group, it was salient that for every unit of increase in the MDD, the odds of making correct judgments decreased to a larger degree in the CE group than the CEK group. When compared to the CE group, the CEK group was more capable of coping with the variance in sentence complexity and displayed a smaller magnitude of influence by the MDD. Based on the fact that the CE and the CEK groups were comparable in the L2 proficiency scores, and the MDD exerted a smaller magnitude of influence on the CEK than on the CE group, it seems plausible to account for the L3 learners' behavior under the SSH. The L2 syntactic constructions with the L3 learners were weaker than those of the native speakers but stronger than those of the L2 learners, residing somewhere in between.

This finding could be incorporated into the framework of the Cumulative Enhancement Model[2]. Being identical in their native language (Mandarin Chinese) and mean L2 scores, the CEK groups were found to have relatively stronger syntactic competence on English reflexive binding than their CE correspondents, being less penetrable to the influence of the syntactic complexity. This could be interpreted in that enhanced experiences of learning an additional language may increase metalinguistic awareness in the L3 group.

In addition, the current finding was in line with the Foreign Language Effect Model[3]. The L2 English and L3 Korean belonged to different types according to the binary division of Huang (1994, 2000a), with English being syntactic-constrained and Korean being pragmatic-constrained. Nevertheless, the CEK group turned out to be more capable of rejecting the influence of syntactic complexity than the CE group, which could be attributed to the status of English and Korean being foreign languages on the one hand, and the contrast of syntactic- and pragmatic orientation on the other hand. When the features of the two foreign languages were not common, it was usually the case that L2 would be converged to reflect the features in L3, as suggested by Aysan (2012).

The CEK group, whose L1 Chinese and L3 Korean both allow for long-distance binding, seemed to be in a substantially advantageous position to the CE group in integrating syntactic constructions with knowledge from other modules or domains. Their grammatical knowledge was relatively more stable regardless of syntactic complexity and achieved higher scores in the TVJT. In this phase, however, we are not able to tease apart the Foreign Language Effect Model and the Cumulative Enhancement Model.

Conclusion

This experiment examined the influence of syntactic complexity on the offline interpretation of L2 reflexive bindings with the L3 learners and compared the data of the L3 group with those of the two L2 groups. The results showed a clear pattern of facilitative influence of L3 on the L2 reflexive resolution. Compared to the CE group, the CEK group was revealed to be more capable of rejecting the effect of an increased sentence complexity and showing a smaller magnitude of variation in their judgment of L2 English reflexive bindings.

Even though the current study found the possible interplay of syntactic and pragmatic constraints on the interpretation of L2 reflexive bindings with L3 learners, some issues remain unresolved. Firstly, the findings of this study do not endow us with sufficient evidence to tease apart the Foreign Language Effect Model and the Cumulative Enhancement Model in accounting for backward transfer with L3 learners. It remains unclear whether the transfer originated from the shared status of being foreign to English and

[2] The Cumulative Enhancement Model stipulates that both the L1 and the L2 can potentially have an influence on L3A, and the influence is positive by nature (Flynn et al., 2004).

[3] The Foreign Language Effect Model, also called the L2 Status Factor, contemplates that the L2 serves as the source of transfer because of its shared foreignness with the L3, and all other subsequently acquired languages (Bardel & Falk, 2007).

Korean, or from the accumulated experience in acquiring more languages. To address this ambiguity, future studies may include a mirror-image L3 group with L1 Korean-L2 English-L2 Chinese language configuration and see if there is a backward transfer and whether the transfer is facilitative or detrimental in nature. Secondly, the stimuli in the Truth Value Judgment Task are not designed to be balanced in the pragmatical congruency factor. Therefore, future studies could manipulate the congruency in animacy, gender, and number in the stimuli to reveal a more comprehensive picture in the case of how the pragmatic constraints modulate the backward transfer with the L3 learners.

Acknowledgement

We are grateful to Mineharu Nakayama for sharing his ideas and valuable inputs. This work was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2019S1A5A2A03038771).

The Authors

Hee-Don Ahn (first author) is professor at Department of English, Konkuk University, Korea.

E-mail: hdahn@konkuk.ac.kr

Yongjoon Cho (co-author*)* is lecturer at Department of Korean, Konkuk University, Korea.

E-mail: ycho@konkuk.ac.kr

Jong-Bai Hwang (co-author) is professor at Department of English Education, Konkuk University, Korea.

E-mail: jongbai@konkuk.ac.kr

Jung Hyun Lim (co-author) is lecturer at Department of English, Konkuk University, Korea.

E-mail: dlightjlim@gmail.com

Cui Mao (corresponding author) is associate professor at School of Foreign languages, Shandong University of Technology, China.

E-mail: 623219460@qq.com

References

Ahn, H. D., & Mao, C., (2019). Reverse transfer of L3 on the interpretation of L2 reflexives. *The Journal of Asia TEFL*, *16*, 1323-1331. http://dx.doi.org/10.18823/asiatefl.2019.16.4.16.1323

Aysan. Z. (2012). *Reverse Interlanguage Transfer: The effects of l3 Italian & l3 French on l2 English pronoun use*. [Unpublished master thesis]. Bilkent University.

Badecker, W., & Straub, K. (2002). The processing role of structural constraints on the interpretation of pronouns and anaphors. *Journal of experimental psychology. Learning, memory, and cognition. 28*, 748-769. https://doi.org/10.1037/0278-7393.28.4.748

Bardel, C., & Falk, Y. (2007). The role of the second language in third language acquisition: The case of Germanic syntax. *Second Language Research*, *23*(4), 459-484.

Clackson, K., Felser, C., & Clahsen, H. (2011). Children's processing of reflexives and pronouns in English: Evidence from eye-movements during listening. *Journal of Memory and Language*, *65*, 128-144.

Clahsen, H., & Felser, C. (2006). Continuity and shallow structures in language processing: A reply to our commentators. *Applied Psycholinguistics*, *27*, 107-126.

Chow, W. Y., & Zhou, Y. (2018). Eye-tracking evidence for active gap-filling regardless of dependency length. *Quarterly Journal of Experimental Psychology*, *72*(6), 1297-1307. https://doi.org/10.1177/1747021818804988

Cunnings, I., & Felser, C. (2013). The role of working memory in the processing of reflexives. *Language and Cognitive Processes*, *28*, 188-219.

Flynn, S., Foley, C., & Vinnitskaya, I. (2004). The Cumulative-Enhancement Model for language acquisition: Comparing adults' and children's patterns of development in first, second and third language acquisition of relative clauses. *The International Journal of Multilingualism, 1*, 3-16.

Gibson, E. (1998). Linguistic complexity: Locality of syntactic dependencies. *Cognition*, *68*, 1-76. https://doi.org/10.1016/S0010-0277(98)00034-1

Huang, Y. (1991). A neo-Gricean pragmatic theory of anaphora. *Journal of Linguistics, 27*(1), 301-335.
Huang, Y. (1994). *The syntax and pragmatics of anaphor*. Cambridge University Press.
Huang, Y. (2000a). *Anaphora: A cross-linguistic study*. Oxford University Press.
Huang, Y. (2000b). Discourse anaphora: Four theoretical models. *Journal of Pragmatics, 32*, 151-176.
Huang, Y. (2004). Anaphora and the pragmatics-syntax interface. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 288-314). Blackwell.
Huang, Y. (2013). Bayesian probabilistic model of discourse anaphoric comprehension, linguistic typology, and neo-Gricean pragmatics. *Theoretical Linguistics, 39*, 95-108.
Huang, Y. (2016). Aspects of anaphora in Chinese and in some Germanic, Romance, and Slavic languages, the 'syntactic' versus 'pragmatic' language typology, and neo-Gricean pragmatics. In K. Allan, A. Capone, & I. Keckes (Eds.), *Pragmemes and theories of language use* (pp. 21-43). Springer.
Hudson, R. (2007). *Language networks: The new word grammar*. Oxford University Press.
Hudson, R. (2010). *An Introduction to word grammar (Cambridge Textbooks in Linguistics)*. Cambridge University Press, Cambridge.
Jiang, J., & Liu, H. (2015). The effects of sentence length on dependency distance, dependency direction and the implications-Based on a parallel English-Chinese dependency treebank. *Language Sciences, 50*, 93-104.
Jiang, J., & Ouyang, J. (2017). Dependency distance: A new perspective on the syntactic development in second language acquisition. *Physics of Life Reviews*, 21, 209-210. https://doi.org/10.1016/j.plrev.2017.06.018
Jiang, Y., & Jiang, X. L. (2019). The influence of the maximal dependency distance on the non-fluency in interpretation. *Foreign Language Research*, *1*(1), 81-88.
Lee, K.-Y. (2008). The role of pragmatics in reflexive interpretation by Korean learners of English. In M. Bowles (Ed.), *Selected proceedings of the 2007 second language research forum* (pp. 97-112). Cascadilla Proceedings Project.
Liu, H. (2008). Dependency distance as a metric of language comprehension difficulty. *Journal of Cognitive Science*, *9*(2), 159-191.
Liu, H., (2010). Dependency direction as a means of word-order typology: A method based on dependency Treebanks. *Lingua 120*(6), 1567-1578. http://dx.doi.org/10.1016/j.lingua.2009.10.001
Liu, H., Xu, C., & Liang, J. (2017). Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews, 21,* 171-193. https://doi.org/10.1016/j.plrev.2017.03.002
Macmillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. Cambridge University Press.
Nivre, J. (2006). *Inductive dependency parsing*. Springer.
Pollard, C., & I. Sag. (1992). Anaphors in English and the scope of binding theory. *Linguistic Inquiry, 23,* 261-303.
Pollard, C., & I. Sag. (1994). *Head-driven phrase structure grammar*. CSLI, University of Chicago Press.
Sperlich, D. (2013). *The acquisition of long-distance reflexives in Chinese as an interlanguage: An experimental study* [Unpublished doctoral dissertation]. The University of Auckland.
Sperlich, D. (2016). Pragmatic or syntactic Ziji? Evidence from language transfer. *Journal of Cognitive Science, 17*(4), 607-652. https://doi/org/10.17791/jcs.2016.17.4.607
Xiang, M., Dillon, B., & Phillips, C. (2009). Illusory licensing effects across dependency types: ERP evidence. *Brain & Language, 108*(1), 40-55. https://doi.org/10.1016/j.bandl.2008.10.002
Yoshimura, N., Nakayama, M., Sawasaki, K., Fujimori, A., & Shimizu, H. (2012). L2 knowledge at the syntax-pragmatics interface: Interpretations of reflexives by Japanese, Korean, and Chinese ESL learners. In Y. Otsu (Ed.), *The proceedings of the 13th Tokyo Conference on Psycholinguistics* (pp. 303-323). Hituzi Syobo.

(Received May 31, 2022; Revised August 20, 2022; Accepted Sep 18, 2022)

