# Influence of Syntactic Knowledge Acqusition and Language Transfer on Dependency of Relative Clauses: Connectionist Simulation*

Masahiro KOBAYASHI

Organization for Educational Support and International Affairs,
Education Center, Tottori University

**Abstract**

Many studies have been done on dependency resolution of the relative clause attachment, especially in the natural language processing field, but there are still a lot of mysteries about how second language language learners acquire dependency resolution. The present study is a computer simulation of the process of Japanese learners of English on the basis of a connectionist network model to investigate how syntactic knowledge influences the learners' performance on the acquisition of the relative clause attachment in English. The simulation results indicate a possibility that the acquisition of the fundamental syntactic knowledge can promote learns' understanding of the choice of the relative clause attachment.

## 1 Introduction

Syntactic ambiguity resolution has been extensively investigated in the area of sentence processing, and it is now widely acknowledged that various types and levels of linguistic knowledge are used to comprehend linguistic input (Hawkins 1994, Frazier 1995, 2013, Juffs and Rodríguez 2015). In addtition, several in-depth research projects on L2 acquisition address the issues of syntactic ambiguity resolution and illustrate how L2 learners resolve the syntactic dependencies, shedding light on how their L1 knowledge influences, or interferes with, the way we process L2 input, which is called L1 transfer. This study uses the three-development phase model proposed by Miyao and Omaki (2006) to investigate to what extent syntactic knowledge can contribute to the acquisition of choice preference of

*This work was supported by JSPS KAKENHI Grant Number JP17K02971.

relative clause attachment to the antecedent noun phrases (NPs) for Japanese L2 Learners of English.

The next section will describe the problems that the current research handles and also review the previous research. Section 3 will give an overview of the computer simulation method based on the connectionist model, and discuss the results of the simulation and some consequences of the experiments. The final section will sum up the discussion and also discuss future research.

## 2 Problems and Previous Studies

This section will delineate the issues that this paper deals with and discuss previous studies. In linguistics, it is well known that there is an ambiguity in the choice of antecedent NPs of the relative clause phrases in constructions such as (1).

(1) Someone hit $[_{\text{NP1}}$ the father of $[_{\text{NP2}}$ the politician$]]$ $[_{\text{RC}}$ who was on the beach$]$.

In (1) the relative clause (RC) "*who was on the beach*" can be attached to NP1, (it is called *High Attachment* ($HA$) because its location in the tree diagram is in the higher position). The relative clause can also modify NP2, where we learn that it is the politician who was on the beach (*Low Attachment*) ($LA$). Furthermore, it has been argued that speakers of particular languages show preferences for High Attachments, whereas others prefer Low Attachments. (Fodor 2002). Such cross-linguistic variation has attracted researchers to investigate how their L1 language knowledge influences the resolution of the relative clause dependencies.

For instance, Frenck-Mestre (2002) used low and high-proficiency English learners of French, and low-proficiency Spanish learners of French to conduct eye-tracing tests. The findings were (i) the low-proficiency English learners of French had an $LA$ preference, which native English speakers do in English, and (ii) the low-proficiency Spanish learners of French showed an $HA$ preference, just like native Spanish speakers do. That research also found that (iii) the high-proficient English learners of French exhibited an $HA$ preference when processing an equivalent construction to (1) in French. Frenck-Mestre (2002) deduced that learners transfer their L1 grammatical knowledge when they interpret the attachment of relative clauses in their L2. Another researcher who argues for L1 transfer in attachment processing is Dussias (2003); testing advanced English learners of Spanish and advanced Spanish learners of English with off-line tests and finding that both groups had an $LA$ preference in their L2. The $LA$ preference of the English learners of Spanish can be accounted for on the basis of transfer theory of grammatical knowledge or parsing strategy, but the fact that the Spanish learners of English preferred the $LA$ interpretation is against the evidence that native Spanish speakers tend to exhibit an $HA$ preference. Dussias (2003) pointed out that they were advanced, high-proficiency learners of English, and had therefore already acquired the processing strategies of their target-language.

In contrast, Papadopoulou and Clahsen (2003) designed on-line and off-line tests, using native Greek speakers as a control group, for Spanish, German, and Russian learners of Greek to investigate their preferences for relative clause attachment, and found that there was no significant difference between any of the groups in L2 processing, although all groups showed an $HA$ preference in their L1. Based on the evidence they found, Papadopoulou and Clahsen (2003) argued that there is no transfer of L1 to the target languages when processing relative clause dependencies in L2. Felser, Roberts, Marinis and Gross (2003) also reported that there seemed to be no transfer effect in interpreting relative clause attachment. Felser et al. (2003) tested high-proficiency Greek and German learners of English and couldn't find any evidence that the participants transferred their L1 grammatical knowledge.

On the other hand, Miyao and Omaki (2006) proposed a three-development phase model, which can be regarded as a compromise between the two opposing groups. Miyao and Omaki (2006) chose intermediate-level Korean learners of Japanese to conduct the on-line and off-line tests. Their off-line test was an interpretation task test where both the Japanese control group and the Korean learners of Japanese demonstrated a clear $HA$ preference beyond the chance-level, which is an indication that there is a transfer effect for L2 learners. But, their on-line test, which was a self-paced reading task test, demonstrated that the Korean participants showed an $LA$ preference in Japanese, even though in their L1 they have an $HA$ preference. To account for these results, Miyao and Omaki (2006) proposed the three-development phase model; the first phase is the stage where learners show full L1 transfer and this is a typical processing strategy observed for low-proficiency learners. The next phase is called the intermediate phase and learners are still in the process of developing their L2 grammar and processing strategy, and in the third stage learners acquire target-like grammar. According to their model, the intermediate-level Korean learners of Japanese in their study are in the second stage so the participants showed a target-like $LA$ preference.

This study will adopt the model proposed by Miyao and Omaki (2006) where L2 learners pass through learning stages, eventually reaching target-like grammar. The study will simulate the learning processes of the Japanese learners of English with several different conditions in order to investigate to what extent grammatical knowledge influences the efficiency and accuracy of acquiring grammatical knowledge of relative clause attachment in L2.

# 3 Connectionist Simulation

This section will briefly illustrate the method of the computer simulation of this paper and discuss the results of the simulation experiments.

## 3.1 Model Description

The current study is meant to model Japanese learners of English. Each simulation run simulates a learner's learning process, and each learner is supposed to go through six stages of the learning process. As Table 1 shows, each run of the simulation represents a learner whose first language is Japanese, so the simulation starts with learning Japanese with simple intransitive and transitive sentences at phase I. The learning data set is generated by a sentence generation program with a dictionary where there are five definite nouns, such as “*Sono Jōyū*” (“the actress”), “*Sono sensei*’ (“the teacher”), and five common nouns, such as “*Meshitsukai*” (“servant”), “*Kodomo*” (“child”), and eight relative clause phrases, such as “*barukoni ni ita*” (“who was on the balcony”), “*kōen ni ita*” (“who was in the park”). There are also six intransitive verbs, such as “*neta*” (“slept”), “*aruita*” (“walked”), and eight transitive verbs, such as “*utta*” (“shot”), and “*oshita*” (“pushed”). The Japanese dictionary includes three postpositions such as, $ga$, $wo$ and a genitive case marker, $no$, on the other hand the English dictionary of this program has a preposition $of$, a copula $was$, and a relative clause pronoun $who$. The sentence generator generates Japanese and English sentences (or phrases), paired with semantic representations of the particular sentences (or phrases) as training data. For instance, when it generates a transitive sentence of Japanese and English, it randomly selects arbitrary nouns and a verb from the word pool to form a sentence, as seen in (2a). It generates a RC attachment target sentence in the same way, as illustrated in (2b).

(2) a. Sono seijika wa sono meshitsukai wo utta .
DEF politician TOP DEF servant ACC shot
“The politician shot the servant.”

b. Sono kashu wa [$_{RC}$ kūkō ni ita ] [[$_{NP2}$ sono sensei no ] $_{NP1}$ otōsan wo ]
DEF singer TOP airport LOC exist DEF teacher GEN father ACC
ketta.
kicked
“The singer kicked the father of the teacher who was in the airport”

Thirty intransitive sentences were generated by randomly combining five definite nouns and six intransitive verbs, and six of these sentences, i.e., one fifth of the generated sentences, were used as testing data. Two hundred transitive sentences were generated by randomly

Table 1: Learning Phases, Learning and Testing Data

| Phase | Learning Data | # of Learning Data | # of Testing Data |
|---|---|---|---|
| I (JPN) | Intrans. & Trans. | 24 + 160 | 6 + 40 |
| II (JPN) | RC Phrases | I + 160 | I + 40 |
| III (JPN) | RC Attachment | II + 160 | II + 40 |
| IV (JPN+ENG) | Intrans. & Trans. | III + 24 + 160 | III + 6 + 40 |
| V (JPN+ENG) | RC Phrases | IV + 160 | IV + 40 |
| VI (JPN+ENG) | RC Attachment | V + 160 | V + 40 |

combining five definite nouns as subjects, eight transitive verbs and five common nouns as objects. One fifth of two hundred transitive sentences (namely, 40 sentences) were used for testing. In phase I, all 184 intransitive and transitive sentences were used for training in a random order, and 46 sentences were used to test the accuracy. This phase simulates the first learning process of simple sentences in L1. At the next stage, 160 RC phrases, such as "*kūkō ni ita sono sensei*" ("the teacher in the airport"), as in (2b), were added to the data set from phase I and used for the training. The testing data consists of 60 RC phrases as well as the data set from phase I. New phase III is where a learner learns the whole data set from phase I and II plus 160 target sentences, such as (2b). The testing data set for phase III includes the dataset from the previous stages as well as forty new target sentences generated in the same way.

In phase IV, a learner starts learning English with a simple data set of intransitive and transitive sentences generated by the sentence generation program in the same way as it generated the Japanese dataset used in phases I to III. A learner learns English RC phrases and the target English sentences in phrases V and VI, respectively.

Labeled data for the training is also produced by the sentence generation program. The learners' task is to learn weights of the network between nodes and biases of each layer to associate the input and the output. Each word from the two languages as well as logical expressions as semantic representations are represented by a 10-unit vector, such as [0, 0, 1, 0, 1, 0, 1, 1, 0, 1]. Each word and its meaning is represented with 10 units, so 80 units out of 86 are used to denote the input sentences or phrases. Two of the remaining six units are an identifier describing whether the input is a sentence or a phrase, and the other two units are used to tell the learner whether an RC expression attaches the higher NP ($HA$) or the lower NP ($LA$) if it contains any RC expression. When the input expression doesn't contain any RC phrases, those units are denoted as "0.5". The last two units are for the language identifier, e.g., the Japanese input is represented as [0, 0] and English [1, 1]. The semantic expression has [0.5, 0.5] as the last two units.

The three-layered neural network is training with a learning rate of 0.004, and the training is repeated 4000 times at each phase. The input layer has 86 units, and the hidden layer consists of 450 units. The output layer has 56 units, and the network is fully connective. The raw output value each unit is calculated with the formula (3), where the value is the sum of the output value $I_i$of the previous layers' unit $i$ for the current unit $j$ multiplied by the weight $W_{ij}$ between these two layers, plus the bias $B_j$ of that unit.

(3) $\sum_{i=1}^{n} W_{ij} I_i + B_j$

The raw value returned by the formula (3) will be moderated by an activation function, and the current study adopts a sigmoid function as the activation function, as defined in (4).

(4) $f(x) = \dfrac{1}{1 + e^{(-x)}}$

The error of the output layer is calculated with the sum-of-square errors, and the error value is back-propagated (Rumelhart, Hinton and Williams 1986) to each layer's units and weights between layers as the differential coefficient ($\partial E$).

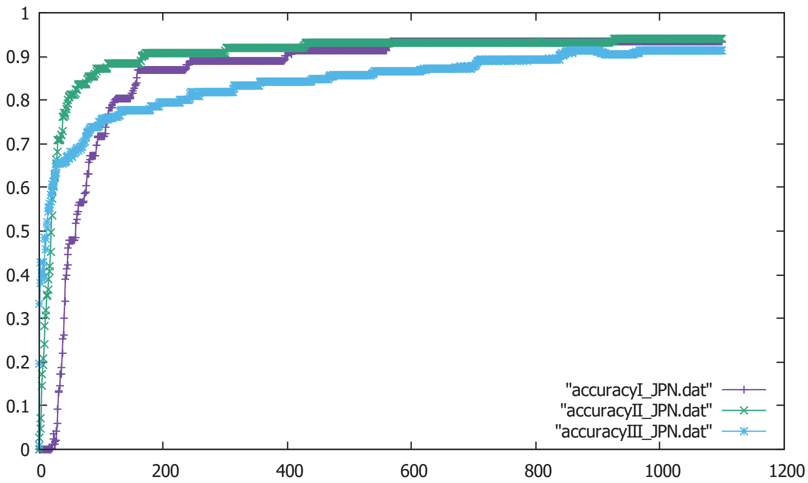


Figure 1: Accuracy of Japanese at Each Phase

## 3.2 Results and Discussion

First, the overall accuracy of each phase was calculated to determine the performance of the network. Figure 1 illustrates the accuracy of a typical run at each phase from I to III, where a learner learns Japanese as outlined in table 1. The average scores of accuracy was calculated every 10 times for printing out the results, so the numbers along the x axis represent one-tenth of the actual epoch numbers. This figure shows that the accuracy of the training reaches 91 to 93% when the simulation is repeated until it stabilizes. The findings, especially at the early stages of each phase, as Figure 2 shows. Figure 2 illustrates the results of the average of five runs of the simulation in the early stages of each phase. As predicted, the first phase takes a longer time to show high accuracy because the learner is learning a language (Japanese) for the first time at this phase. In phase II, the learner learns RC phrases, but it had already obtained knowledge of intransitive and transitive constructions before it reached that phase, so it took a shorter time to show over 90% performance. On the other hand, the average accuracy of the last phase shows that it traced a similar route to phase II, but after it achieved 60%, it took longer to get the 90% accuracy or above, because it has to handle long sentence, such as (2b).

Figure 3 shows the accuracy of English expressions. The test data includes Japanese expressions as well as English to assess the accuracy. The tendency of the accuracy transition is similar to that of Japanese in the sense that phase IV took longer to achieve 70% accuracy, but even though the final phase VI reached 70% as quickly as phase V did, it eventually took longer time to achieve high accuracy.

The next simulation was conducted to determine how syntactic knowledge of L2 influences the accuracy of the L2. In the English learning stages (Phase IV to VI) of the simulation, intransitive and transitive simple sentences were used to train the network in

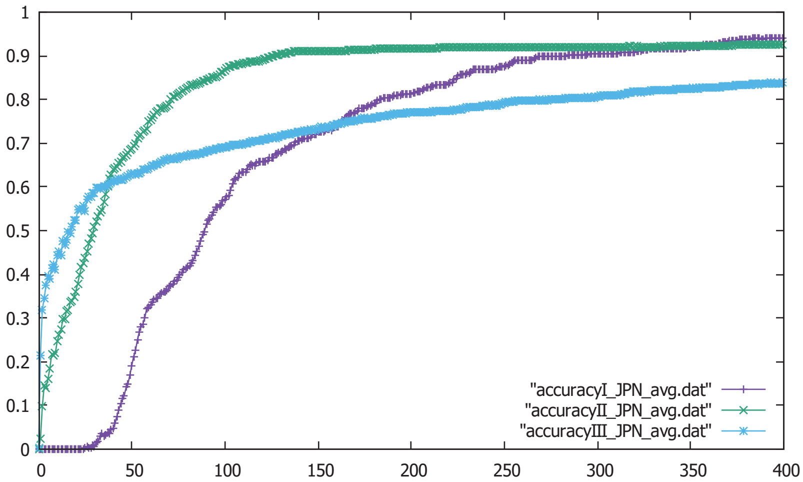


Figure 2: Average Scores of Accuracy at Early Stages

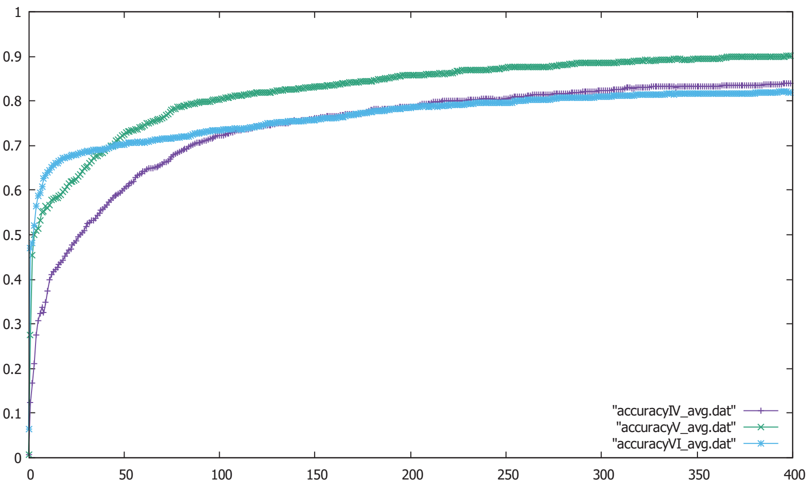


Figure 3: Average Scores of Accuracy of Phase IV to VI

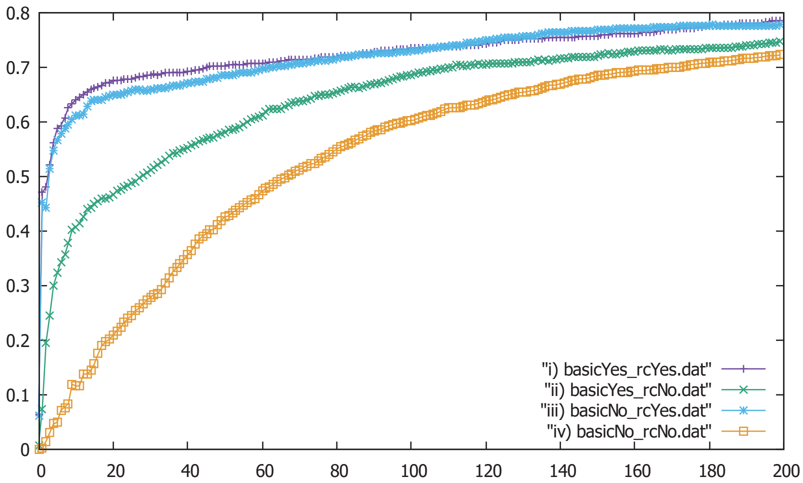


Figure 4: Average Scores of Accuracy of Four Conditions

phase IV, and the RC phrases were added to the data set in the next phase (Phase V). The last phase VI simulates how L2 learners learn the RC attachment. The two parameters, the learning of the intransitive / transitive constructions and learning the RC phrases, were used to investigate how syntactic knowledge works for L2 learning. There are four learning situations as a result of the combination of the two parameters; (i) L2 learners learn the RC phrases as well as simple intransitive and transitive constructions as basic syntactic knowledge in phases IV and V (Basic-Yes-RC-Yes type), or (ii) the network learns the intransitive and transitive sentences in phase IV, and it keeps using the same intransitive and transitive sentences but not learning the RC phrases until phase VI (Baisc-Yes-RC-No type), or (iii) the network learns only the RC phrases in phase V without learning the intransitive and transitive sentences as a basis of English in phase IV (Baisc-No-RC-Yes type). Finally, (iv) the network does not learn English at all, but it suddenly starts learning the RC attachment in the final phase VI (Baisc-No-RC-No type). Figure 4 shows the accuracy average of five runs of the simulation, and the performances of each network is assessed at the final phase VI with only the target sentence test data. As expected, condition (iv) shows the lowest performance at any stage of the learning process because the network didn't have any knowledge of the simple constructions of English and RC structure before it learned the RC attachment preference. What is intriguing here is that the network showed a slightly better performance under condition (i) than condition (iii), and also showed better score under condition (ii) than condition (iv) especially in the early stages. Contrary to the results, the prediction was that learning English intransitive and transitive sentences would be "noise" for machine learning of the attachment preference, because the RC phrases occupy six slots (60 units) out of the entire target sentence (80 units), which accounts for 75% of the slots. These results, however, imply that there is a possibility that the network obtained some basic understanding about English vocabulary, and basic word order before it started learning the RC attachment preference, and this kind of syntactic knowledge helped the network achieve better performance in the early stages.

## 4 Conclusion

This paper investigated how syntactic knowledge can influence the accuracy performance of the attachment resolution of relative clause constructions. The study adopted the idea, proposed by Miyao and Omaki (2006), that there are three stages, from the low-proficiency speaker to the high-proficiency speaker, in terms of L1 transfer. First, the overall performance of the network was assessed through six phases and found that phase I showed relatively weaker performance than other phases due to a lack of the knowledge of the target language, and phase III showed good performance at early stages, but it took a longer time to achieve the same level of accuracy, perhaps because of the wider coverage of the data set, in both Japanese and English. To assess the performance of the acquisition of the attachment preference of English as L2, the simulation was conducted with four different conditions, and the results implied that there is a possibility that if learners have access to fundamental grammatical knowledge, such as word order of etc., it may allow learners to show better performance, especially at earlier stages of the learning process.

For future research, I will simulate other possible syntactic factors which can promote learner's better understanding or acquisition of the modification dependency resolution. Future research should also explore other network models, such as the Recurrent Neural Networks, to simulate on-line tasks. I will also test Japanese language learners as participants to build a human model, and will compare the similarity and difference between the the models to provide a feedback for the model in this research.

## References

Dussias, P. E. 2003. "Syntactic Ambiguity Resolution in L2 Learners: Some Effects of Bilinguality on L1 and L2 Processing Strategies." *Studies in Second Language Acquisition.* 25. pp.529-557.

Felser, C., Roberts, L., Marinis, T., and Gross, R. 2003. "The Processing of Ambiguous Sentences by First and Second Language learners of English." *Applied Psycholinguistics.* 24. pp.453-489.

Frazier, L. 1995. "Constraint Satisfaction as a Theory of Sentence Processing." *Journal of Psycholinguistic Research.* 24(6). pp.437-468.

Frazier, L. 2013. "Syntax in Sentence Processing." In R. P. G. van Gompel (ed.) *Sentence Processing.* pp.21-50. Psychology Press.

Frenck-Mestre, C. 2002. “An On-line Look at Sentence Processing in the Second Language.” In Heredia, R. and Altarriba, J. (eds.) *Bilingual Sentence Processing*. pp.217-236.

Fodor, J. D. 2002. “Prosodic Disambiguation in Silent Reading: An On-line Look.” In Hirotani, M. (ed.) *NELS 32*. pp.113-132.

Hawkins, A., J. 1994. *A Performance Theory of Order and Constituency*. Cambridge University Press.

Juffs, A. and Rodríguez, G. 2015. *Second Language Sentence Processing*. Routledge.

Miyao, M. and Omaki, A. 2006. “No Ambiguity about it: Korean Learners of Japanese Have a Clear Attachment Preference.” In Bamman, D, Magnitskaia, T., and Zaller, C. (eds.) *Proceedings of the 30th Annual Boston University Conference on Language Development Supplement*. pp.1-12.

Omaki, A. and Schulz, B. 2011. “Filler-gap Dependencies and Island Constraints in Second-language Sentence Processing.” *Studies in Second Language Acquisition*. 33. pp.563-588.

Papadopoulou, D. and Clahsen, H. 2003. “Parsing Strategies in L1 and L2 Sentence Processing: A Study of Relative Clause Attachment in Greek.” *Studies in Second Language Acquisition*. 25. pp.501-528.

Rumelhart, D., G., Hinton and R., Williams. 1986. “Learning representations by back-propagating errors.” *Nature*. 323. pp.533-536.